



Catalogue 71-526

Methodology of the Canadian Labour Force Survey

1984-1990

Catalogue 71-526

Méthodologie de l'enquête sur la population active du Canada

1984-1990

Gouvernement
Publications

71-526



Statistics
Canada

Statistique
Canada

Canada

Data in Many Forms . . .

Statistics Canada disseminates data in a variety of forms. In addition to publications, both standard and special tabulations are offered on computer print-outs, microfiche and microfilm, and magnetic tapes. Maps and other geographic reference materials are available for some types of data. Direct access to aggregated information is possible through CANSIM, Statistics Canada's machine-readable data base and retrieval system.

How to Obtain More Information

Inquiries about this publication and related statistics or services should be directed to:

Labour Force Survey Methods Section,
Social Survey Methods Division,

Statistics Canada, Ottawa, K1A 0T6 (Telephone: 951-9893) or to the Statistics Canada reference centre in:

St. John's	(772-4073)	Winnipeg	(983-4020)
Halifax	(426-5331)	Regina	(780-5405)
Montreal	(283-5725)	Edmonton	(495-3027)
Ottawa	(951-8116)	Calgary	(292-6717)
Toronto	(973-6586)	Vancouver	(666-3691)

Toll-free access is provided in all provinces and territories, **for users who reside outside the local dialing area** of any of the regional reference centres.

Newfoundland and Labrador	1-800-563-4255
Nova Scotia, New Brunswick and Prince Edward Island	1-800-565-7192
Quebec	1-800-361-2831
Ontario	1-800-263-1136
Manitoba	1-800-542-3404
Saskatchewan	1-800-667-7164
Alberta	1-800-282-3907
Southern Alberta	1-800-472-9708
British Columbia (South and Central)	1-800-663-1551
Yukon and Northern B.C. (area served by NorthwesTel Inc.)	Zenith 0-8913
Northwest Territories (area served by NorthwesTel Inc.)	Call collect 403-495-2011

How to Order Publications

This and other Statistics Canada publications may be purchased from local authorized agents and other community bookstores, through the local Statistics Canada offices, or by mail order to Publication Sales, Statistics Canada, Ottawa, K1A 0T6.

1(613)951-7277

Facsimile Number 1(613)951-1584

National toll free order line 1-800-267-6677

Toronto
Credit card only (973-8018)

Des données sous plusieurs formes . . .

Statistique Canada diffuse les données sous formes diverses. Outre les publications, des totalisations habituelles et spéciales sont offertes sur imprimés d'ordinateur, sur microfiches et microfilms et sur bandes magnétiques. Des cartes et d'autres documents de référence géographiques sont disponibles pour certaines sortes de données. L'accès direct à des données agrégées est possible par le truchement de CANSIM, la base de données ordiolinguée et le système d'extraction de Statistique Canada.

Comment obtenir d'autres renseignements

Toutes demandes de renseignements au sujet de cette publication ou de statistiques et services connexes doivent être adressées à:

Section des méthodes de l'enquête sur la population active,
Division des méthodes d'enquêtes sociales,

Statistique Canada, Ottawa, K1A 0T6 (téléphone: 951-9893) ou au centre de consultation de Statistique Canada à:

St. John's	(772-4073)	Winnipeg	(983-4020)
Halifax	(426-5331)	Regina	(780-5405)
Montréal	(283-5725)	Edmonton	(495-3027)
Ottawa	(951-8116)	Calgary	(292-6717)
Toronto	(973-6586)	Vancouver	(666-3691)

Un service d'appel interurbain sans frais est offert, dans toutes les provinces et dans les territoires, **aux utilisateurs qui habitent à l'extérieur des zones de communication locale** des centres régionaux de consultation.

Terre-Neuve et Labrador	1-800-563-4255
Nouvelle-Écosse, Nouveau-Brunswick et Île-du-Prince-Édouard	1-800-565-7192
Québec	1-800-361-2831
Ontario	1-800-263-1136
Manitoba	1-800-542-3404
Saskatchewan	1-800-667-7164
Alberta	1-800-282-3907
Sud de l'Alberta	1-800-472-9708
Colombie-Britannique (sud et centrale)	1-800-663-1551
Yukon et nord de la C.-B. (territoire desservi par la NorthwesTel Inc.)	Zénith 0-8913
Territoires du Nord-Ouest (territoire desservi par la NorthwesTel Inc.)	Appelez à frais virés au 403-495-2011

Comment commander les publications

On peut se procurer cette publication et les autres publications de Statistique Canada auprès des agents autorisés et des autres librairies locales, par l'entremise des bureaux locaux de Statistique Canada, ou en écrivant à la Section des ventes des publications, Statistique Canada, Ottawa, K1A 0T6.

1(613)951-7277

Numéro du télégraphe 1(613)951-1584

Commandes: 1-800-267-6677 (sans frais partout au Canada)

Toronto
Carte de crédit seulement (973-8018)

Statistics Canada
Social Survey Methods Division

Statistique Canada
Division des méthodes d'enquêtes sociales

Methodology of the Canadian labour force survey

Méthodologie de l'enquête sur la population active du Canada

1984-1990

1984-1990

M.P. Singh
J.D. Drew
J.G. Gambino
F. Mayda

M.P. Singh
J.D. Drew
J.G. Gambino
F. Mayda

Published under the authority of the Minister
of Industry, Science and Technology

© Minister of Supply
and Services Canada 1990

All rights reserved. No part of this publication
may be reproduced, stored in a retrieval system or
transmitted in any form or by any means, electronic,
mechanical, photocopying, recording or otherwise
without prior written permission of the Minister of
Supply and Services Canada.

December 1990

Price: Canada: \$48.00
United States: US \$58.00
Other Countries: US \$67.00

Catalogue 71-526

ISBN 0-660-54217-X

Ottawa

Publication autorisée par le ministre de
l'Industrie, des Sciences et de la Technologie

© Ministre des Approvisionnements
et Services Canada 1990

Tous droits réservés. Il est interdit de reproduire ou de
transmettre le contenu de la présente publication, sous quelque
forme ou par quelque moyen que ce soit, enregistrement sur
support magnétique, reproduction électronique, mécanique,
photographique, ou autre, ou de l'emmagasiner dans un système
de recouvrement, sans l'autorisation écrite préalable du
ministre des Approvisionnements et Services Canada.

Décembre 1990

Prix : Canada : 48 \$
États-Unis : 58 \$ US
Autres pays : 67 \$ US

Catalogue 71-526

ISBN 0-660-54217-X

Ottawa

CAI
BS71
-1990
C526

Acknowledgement

Many parts of Statistics Canada's organization are involved in the functioning of the Labour Force Survey. Divisions or Branches with a major involvement in the survey include the following: Household Surveys Division, responsible for management of the survey, data dissemination and user liaison; Labour and Household Surveys Analysis Division, responsible for the survey content, questionnaire, and for data development and analysis; Survey Operations Branch, responsible for field operations, and data capture and processing in Regional Offices; Methodology Branch, responsible for the sample design, data collection methodology, estimation procedures, and quality evaluation; and Informatics Branch, responsible for electronic data processing services to both Ottawa and the Regional Offices.

In 1980, a Redesign Committee was established with members from the above areas, along with members from other areas whose programs are associated with the LFS infrastructure. The Redesign Committee met until 1985 to provide direction and guidance on the development and implementation of the sample redesign.

The Redesign Committee was chaired by M.P. Singh. It benefited from the active participation of members of senior Statistics Canada management, namely I.P. Fellegi, D.B. Petrie and G.J. Brackstone. Other members from the principal participating areas included K. Bennett, M. Brochu, D. Drew, M. Levine, I. Macredie, R. Platek and R. Ryan.

This report was written by M.P. Singh, D. Drew, J. Gambino and F. Mayda, with help in preparation and review of earlier drafts from K. Bennett, H. Choudhry, G.B. Gray and J. Lindeyer. The document was translated by L. Macklovitch, style edited by C. Campbell and processed by J. Clarke, C. Larabie and C. Lacroix.

Remerciements

La réalisation de l'enquête sur la population active fait intervenir un grand nombre de services de Statistique Canada. Les divisions ou directions suivantes jouent un rôle de premier plan: la Division des enquêtes-ménages, qui est responsable de la gestion de l'enquête, de la diffusion des données et de la liaison avec les utilisateurs; la Division de l'analyse des enquêtes sur le travail et auprès des ménages, qui est responsable du contenu de l'enquête, du questionnaire et de l'élaboration et de l'analyse des données; la Direction des opérations des enquêtes, qui est chargée des opérations sur le terrain et de la saisie et du traitement des données dans les bureaux régionaux; la Direction de la méthodologie, qui est responsable du plan d'échantillonnage, des méthodes de collecte des données, des procédures d'estimation et de l'évaluation de la qualité; la Direction de l'informatique, qui est responsable des services informatiques de traitement des données à Ottawa et dans les bureaux régionaux.

En 1980 était mis sur pied un Comité du remaniement composé de représentants des services susmentionnés de même que des services dont les programmes sont liés à l'infrastructure de l'EPA. Ce comité, qui s'est réuni jusqu'en 1985, a orienté et guidé l'élaboration et la mise en oeuvre du remaniement de l'échantillon.

Le Comité du remaniement était présidé par M.P. Singh, qui pouvait compter sur la participation active de membres de la haute direction de Statistique Canada, soit I.P. Fellegi, D.B. Petrie et G.J. Brackstone. Les principaux services participants étaient représentés par K. Bennett, M. Brochu, D. Drew, M. Levine, I. Macredie, R. Platek et R. Ryan.

La présente publication a été rédigée par M.P. Singh, D. Drew, J. Gambino et F. Mayda. Par ailleurs, K. Bennett, H. Choudhry, G.B. Gray et J. Lindeyer ont apporté leur concours à la préparation et à la révision des premières ébauches. Le document a été traduit par L. Macklovitch, le style a été revu par C. Campbell et les textes traités par J. Clarke, C. Larabie et C. Lacroix.



Foreword

Since its inception in 1948, the Labour Force Survey (LFS) has undergone a sample redesign following each decennial Census of Population. The purposes of these redesigns include updating of the sample design to reflect changes in the size and characteristics of the population; adoption of improved sampling and estimation methods; and achievement of efficiencies through increased automation in the sample design and survey operations. Redesigns also provide an opportunity to respond to changes in the information needs to be met by the survey.

The redesign of the LFS during the early 1980s responded to the demand for better subprovincial data, the need to identify efficiencies within the LFS, and the need for flexibility in the sample design. Efficiency gains were then redirected to new initiatives, such as the General Social Survey, and to programs which were inadequately funded. Flexibility in the sample design permitted future modifications to sample sizes when information needs and budgetary circumstances changed.

Improvements in the cost efficiency of the survey design included reduction in the clustering of the sample, better stratification and sample allocation schemes, and more telephone interviewing. The combined effect of these improvements permitted a reduction in the overall sample size of 7%, lower per household data collection costs, and a significant reduction in the sampling error for the smaller subprovincial regions, without an increase in sampling error for provincial and national estimates.

The flexibility built into the design for sample size changes was needed in 1989-90, when a major sample size increase was implemented to improve the reliability of survey estimates for Unemployment Insurance Commission Regions.

This report describes the methodology introduced in 1985, when the sample redesign based on the 1981 Census was implemented. Also included are changes that took place as late as 1990.

Predecessors to this report are Statistics Canada, Catalogue 71-526, which describes the LFS methodology for the period 1975-84, and Statistics Canada, Catalogue 71-504, which describes the LFS methodology for the period 1965-1974.

Statistics Canada welcomes comments and suggestions from any users of this report regarding its usefulness and how it may be improved in the future.

Ivan P. Fellegi
Chief Statistician of Canada

Avant-propos

L'enquête sur la population active (EPA) existe depuis 1948 et son échantillon a toujours été remanié après chaque recensement décennal de la population. Par ces remaniements, on cherchait notamment à mettre à jour le plan d'échantillonnage en fonction des changements dans la taille et les caractéristiques de la population, à adopter de meilleures méthodes d'échantillonnage et d'estimation et à accroître le rendement par une plus grande automatisation du plan d'échantillonnage et des opérations d'enquête. Les remaniements étaient aussi pour Statistique Canada l'occasion de répondre aux nouveaux besoins en données d'enquête.

Le remaniement de l'EPA au début des années 1980 avait pour but la production de meilleures données infraprovinciales en réponse à des besoins exprimés, la réalisation d'économies grâce à un meilleur rendement et l'établissement d'un plan d'échantillonnage offrant une plus grande marge de manœuvre. Les ressources économisées dans le cadre de l'EPA furent réaffectées à d'autres enquêtes, notamment à l'enquête sociale générale, et à des programmes insuffisamment financés. La moins grande rigidité du plan d'échantillonnage donnait à Statistique Canada la possibilité de modifier la taille des échantillons lorsque les besoins en information et les circonstances budgétaires changeaient.

Les améliorations apportées au plan d'enquête pour en accroître l'efficacité comprenaient la réduction du nombre de grappes dans l'échantillon, de meilleures méthodes pour la stratification et la répartition de l'échantillon et davantage d'interviews téléphoniques. Toutes ces mesures réunies se traduisirent par une réduction de 7 % de la taille globale de l'échantillon, une diminution des coûts de collecte par ménage et un abaissement important de l'erreur d'échantillonnage pour les petites régions infraprovinciales, et ce sans détérioration des estimations provinciales et nationales.

La marge de manœuvre qu'offrait le plan d'échantillonnage au niveau de la taille de l'échantillon fut fort utile en 1989-1990 lorsqu'il fut décidé d'augmenter considérablement la taille de l'échantillon pour accroître la fiabilité des estimations d'enquête pour les régions définies aux fins de l'assurance-chômage.

La présente publication décrit la méthodologie adoptée en 1985, année de la mise en œuvre de l'échantillon remanié en fonction du recensement de 1981. De plus, les changements apportés par la suite, qui dans certains cas ne remontent qu'à 1990, y sont aussi été intégrés.

Statistique Canada a diffusé sur le même sujet deux autres publications, la première sur la méthodologie de l'EPA pour la période 1975-1984 (n° 71-526 au catalogue) et la seconde, sur la méthodologie de l'EPA pour la période 1965-1974 (n° 71-504 au catalogue).

Tous les utilisateurs sont invités à faire parvenir à Statistique Canada leurs commentaires sur l'utilité de la présente publication et leurs suggestions sur la façon de l'améliorer.

Ivan P. Fellegi
Statisticien en chef du Canada

Table of Contents

	Page
Introduction	
1.1 Estimates from the Survey	9
1.2 An Overview of the Survey	10
1.3 Sample Design Prior to 1985	11
1.4 Objectives of the Redesign	12
1.5 Scope of the Report	14
Stratification	
2.1 Introduction	17
2.2 Stratification Algorithm	19
2.3 Stratification in Self Representing Areas	21
2.4 Stratification and Delineation of PSUs in Non Self Representing Areas	23
2.4.1 General Considerations	23
2.4.2 Urban-Rural Design	25
2.4.3 Mixed Design	26
Self Representing Areas	
3.1 Introduction	27
3.2 Formation of Self Representing Units	27
3.3 Sample Design for Area Frame	27
3.3.1 Choice of Clusters	27
3.3.2 Formation of Clusters	29
3.3.3 Choice of the Selection Method	31
3.3.4 Sampling Methodology	32
3.4 Sample Update Methodology	33
3.5 Sample Design for the Apartment Frame	34
Non Self Representing Areas	
4.1 Introduction	37
4.2 Rationale for Changes in the Design	37
4.2.1 Extension of Telephone Interviewing	37
4.2.2 Elimination of Stage of Sampling in Rural Areas	38
4.2.3 Design with Urban/Rural Stratification	39
4.2.4 Special Design for Prince Edward Island	39
4.3 Selection of Primary Sampling Units	39
4.4 Sub-Sampling in Rural Strata	41
4.4.1 Sampling within Primary Sampling Units	41
4.4.2 Selection of Dwellings	42
4.5 Sub-Sampling in Urban Strata	42
4.5.1 Clustering of Urban Centres	42
4.5.2 Selection of Clusters	42
4.6 Sub-Sampling in Mixed Strata	43
4.7 Sample Design for Special Areas	44
4.7.1 Rationale for Special Areas	44
4.7.2 Design of the Remote Areas	44
4.7.3 Design of the Institution Frame	45

Table des matières

	Page
Introduction	
1.1 Estimations produites d'après les résultats de l'enquête sur la population active.	9
1.2 Aperçu général de l'enquête	10
1.3 Plan de sondage avant 1985	11
1.4 Objectifs du remaniement	12
1.5 Objet de la publication	14
Stratification	
2.1 Introduction	17
2.2 Algorithme de stratification	19
2.3 Stratification dans les secteurs autoreprésentatifs	21
2.4 Stratification et délimitation des UPE dans les secteurs non autoreprésentatifs	23
2.4.1 Critères généraux	23
2.4.2 Plan urbain-rural	25
2.4.3 Plan mixte	26
Secteurs autoreprésentatifs	
3.1 Introduction	27
3.2 Formation des unités autoreprésentatives	27
3.3 Plan de sondage pour la base aréolaire	27
3.3.1 Choix des grappes	27
3.3.2 Formation de grappes	29
3.3.3 Choix de la méthode de sélection	31
3.3.4 Méthodes d'échantillonnage	32
3.4 Méthode de mise à jour de l'échantillon	33
3.5 Plan de sondage pour la base d'appartements	34
Unités non autoreprésentatives	
4.1 Introduction	37
4.2 Raison des changements apportés au plan de sondage	37
4.2.1 Interviews téléphoniques	37
4.2.2 Élimination d'un degré d'échantillonnage dans les régions rurales	38
4.2.3 Plan avec stratification urbaine/rurale	39
4.2.4 Plan spécial pour l'Île-du-Prince-Édouard	39
4.3 Sélection des unités primaires d'échantillonnage	39
4.4 Sous-échantillonnage dans les strates rurales	41
4.4.1 Échantillonnage dans les UPE	41
4.4.2 Sélection des logements	42
4.5 Sous-échantillonnage dans les strates urbaines	42
4.5.1 Formation des grappes dans les centres urbains	42
4.5.2 Sélection des grappes	42
4.6 Sous-échantillonnage dans les strates mixtes	43
4.7 Plan de sondage pour les secteurs spéciaux	44
4.7.1 Raison d'être des secteurs spéciaux	44
4.7.2 Plan de sondage pour les régions éloignées	44
4.7.3 Conception de la base d'établissements institutionnels	45

Table of Contents – Continued

	Page
Sample Size Determination	
5.1 Introduction	47
5.2 Data Reliability Objectives	48
5.3 Steps in Allocating the Sample	49
5.3.1 Optimum Relative Sampling Rates	49
5.3.2 Adjustments to the Optimum Allocation	51
5.3.3 Impact of Sample Size Reductions	53
5.3.4 Sample Size Increase in 1989-90	54
Sample Rotation	
6.1 Introduction	55
6.2 Method of Rotation	55
6.3 Operational Steps Involved in Sample Rotation	57
6.4 Assigning Rotation Numbers	58
Survey Operations	
7.1 Introduction	61
7.2 Organization and Responsibility	61
7.3 Survey Schedule	63
7.4 Sampling Activities	64
7.5 Maintenance	66
7.6 Data Collection and Capture Activities	68
7.7 Data Processing	70
7.8 Dissemination of Data	73
Weighting and Estimation	
8.1 Introduction	75
8.2 The Basic Weight	76
8.3 Rural-Urban Factor	76
8.4 Treatment of Non response	77
8.5 Cluster Subweight	78
8.6 Stabilization Factor	78
8.7 Estimation Methodology	78
8.8 Algebraic Description of Weighting	80
8.9 An Example of Weighting a Record	82
8.10 Variance Estimation	84
Data Quality	
9.1 Introduction	85
9.2 Sampling Error	85
9.3 Presentation of Sampling Errors	87
9.4 Non-sampling Errors	88
9.5 Data Collection Quality	92
9.6 Operations and Analysis	94
General Use of the Labour Force Survey Vehicle	
10.1 Introduction	97
10.2 Special Surveys	97
10.3 Supplementary Surveys	98
10.3.1 Examples of Supplementary Surveys	98
10.3.2 Sample Selection Options	98
10.3.3 Data Collection Methodologies	99
10.3.4 Acceptability Criteria for Supplementary Surveys	99
10.4 Examples	99

Table des matières – suite


	Page
Détermination de la taille de l'échantillon	
5.1 Introduction	47
5.2 Objectifs de fiabilité des données	48
5.3 Répartition de l'échantillon	49
5.3.1 Fractions de sondage relatives optimales	49
5.3.2 Modifications apportées à la répartition optimale	51
5.3.3 Conséquences de la réduction de la taille de l'échantillon	53
5.3.4 Augmentation de la taille de l'échantillon en 1989-1990	54
Renouvellement de l'échantillon	
6.1 Introduction	55
6.2 Méthode de renouvellement	55
6.3 Processus de renouvellement de l'échantillon	57
6.4 Attribution des numéros de renouvellement	58
Organisation et responsabilités	
7.1 Introduction	61
7.2 Organisation et Responsabilités	61
7.3 Calendrier de l'enquête	63
7.4 Activités d'échantillonnage	64
7.5 Mise à jour	66
7.6 Collecte et saisie des données	68
7.7 Traitement des données	70
7.8 Diffusion des données	73
Pondération et estimation	
8.1 Introduction	75
8.2 Poids de base	76
8.3 Facteur rural-urbain	76
8.4 Traitement de la non-réponse	77
8.5 Sous-poids de grappe	78
8.6 Facteur de stabilisation	78
8.7 Méthode d'estimation	78
8.8 Description algébrique de la pondération	80
8.9 Comment pondérer un enregistrement: un exemple	82
8.10 Estimation de la variance	84
Qualité des données	
9.1 Introduction	85
9.2 Erreur d'échantillonnage	85
9.3 Présentation des erreurs d'échantillonnage	87
9.4 Erreurs non dues à l'échantillonnage	88
9.5 Qualité de la collecte des données	92
9.6 Opérations et analyse	94
Utilisation de l'enquête sur la population active pour la réalisation d'autres enquêtes	
10.1 Introduction	97
10.2 Enquêtes spéciales	97
10.3 Enquêtes supplémentaires	98
10.3.1 Exemples d'enquêtes supplémentaires	98
10.3.2 Possibilités de plans de sondage	98
10.3.3 Méthodes de collecte des données	99
10.3.4 Critères d'admissibilité relatifs aux enquêtes supplémentaires	99
10.4 Exemples	99

Table of Contents – Concluded

	Page
Developmental Projects	
11.1 Introduction	101
11.2 Small Area Estimation	102
11.3 Composite Estimation	103
11.4 Gross Flows	105
11.5 Longitudinal Data	107
11.6 Variance Studies	108
11.7 Telephone and CATI Development	108
11.8 Address Register	110
Appendixes	
A Numbering System	113
B Sample Size Stabilization	121
C Sample Phase-in	131
D Diagram of the LFS Frame	139
E Composition of the Sample	141
F Forms	167
G Definitions	173
References	175

Table des matières – fin

	Page
Projets de développement	
11.1 Introduction	101
11.2 Estimation pour les petites régions	102
11.3 Estimation composite	103
11.4 Flux bruts	105
11.5 Données longitudinales	107
11.6 Études de la variance	108
11.7 Nouvelles façons d'utiliser l'interview téléphonique et l'interview téléphonique assistée par ordinateur	108
11.8 Registre d'adresses	110
Annexes	
A Système de numérotation	113
B Stabilisation de la taille de l'échantillon	121
C Introduction de l'échantillon	131
D Diagramme de la base de l'EPA	139
E Composition de l'échantillon	141
F Formules	167
G Définitions	173
Bibliographie	175



Digitized by the Internet Archive
in 2023 with funding from
University of Toronto

<https://archive.org/details/39251716070010>

Chapter 1

Introduction

1.1 Estimates from the Survey

The Canadian Labour Force Survey (LFS) was developed following the Second World War to satisfy a need for reliable and timely data on the labour market. Information was urgently required on the massive labour market changes involved in the transition from a war-time to a peace-time economy. The survey was designed to provide estimates of employment by industry and occupation at the regional as well as the national level.

A quarterly survey initially, the LFS became a monthly survey in 1952. In 1960, the Interdepartmental Committee on Unemployment Statistics recommended that the LFS be designated the source of the official measure of unemployment in Canada. This endorsement was followed by a demand for a broader range of labour market statistics, in particular more detailed regional data. The information generated by the survey has expanded considerably over the years and now provides a rich and detailed picture of the Canadian labour market.

The LFS is the only source of monthly estimates of total employment (both full and part time), of self-employment (both agricultural and non-agricultural) and of total unemployment. It publishes monthly standard labour market indicators such as the unemployment rate, the employment-to-population ratio, and the participation rate. It is a major source of information on the personal characteristics of the working-age population, including age, marital status, educational attainment, and family characteristics.

Employment estimates include detailed breakdowns by industry and occupation, job tenure, and usual and actual hours worked. The survey incorporates questions permitting analyses of many topical issues, such as involuntary part-time employment, multiple job-holding, and absence from work.

Unemployment estimates are produced by industry and occupation as well as by duration of unemployment, type of work sought, and activity before looking for work. Some information is also available on the recent labour market involvement (i.e. within the last five years) of persons not in the labour force.

In addition to providing national and provincial estimates, the Labour Force Survey also releases estimates for subprovincial areas, such as Economic Regions and Census Metropolitan Areas. In recent years, estimates of the standard labour market

Chapitre 1

Introduction

1.1 Estimations produites d'après les résultats de l'enquête sur la population active

L'enquête sur la population active (EPA) a été créée après la Deuxième Guerre mondiale suite à un besoin immédiat de données fiables et actuelles portant sur le marché du travail et reflétant le passage d'une économie de guerre à une économie de paix. L'enquête a été conçue dans le but de produire des estimations de l'emploi selon la branche d'activité et la profession tant au niveau régional que national.

Trimestrielle à l'origine, l'enquête sur la population active est devenue une enquête mensuelle en 1952. En 1960, le Comité interministériel de la statistique du chômage a recommandé qu'elle devienne l'instrument officiel par lequel le chômage au Canada serait mesuré. Après l'adoption de cette recommandation, la demande de données a augmenté : les utilisateurs voulaient disposer d'une plus vaste gamme de statistiques sur le marché du travail et en particulier de données régionales plus détaillées. L'éventail d'estimations produites au moyen des résultats de l'enquête s'est élargi considérablement au fil des ans, et représente aujourd'hui un portrait étoffé et détaillé du marché du travail canadien.

L'enquête est la seule source d'estimations mensuelles de l'emploi total, à plein temps et à temps partiel, de l'emploi autonome, dans le secteur agricole et non agricole, et du chômage total. À partir de ces résultats, on publie tous les mois des indicateurs standard du marché du travail comme le taux de chômage, le ratio de l'emploi à la population et le taux d'activité. Il s'agit d'une des principales sources d'information sur les caractéristiques personnelles de la population en âge de travailler (dont l'âge, l'état matrimonial, le niveau d'instruction et la situation familiale).

Les estimations de l'emploi sont ventilées à un niveau détaillé selon, notamment, la branche d'activité, la profession, la durée d'emploi, le nombre d'heures habituellement travaillées et le nombre d'heures réellement travaillées. Certaines des questions posées permettent d'étudier une grande variété de sujets d'actualité comme le travail à temps partiel non choisi, le cumul d'emplois et l'absentéisme.

Les estimations du chômage sont produites par branche d'activité et par profession ainsi que selon la durée du chômage, le genre de travail recherché et l'activité avant la recherche de travail. Il existe également des données sur l'activité récente sur le marché du travail (c.-à-d. au cours des cinq dernières années) des personnes actuellement inactives.

En plus des estimations nationales et provinciales, on produit également des estimations pour des régions infra-provinciales comme les régions économiques et les régions métropolitaines de recensement. On a également eu recours, ces dernières années, à des techniques d'estimation spéciales

indicators have also been tabulated for small areas, such as Census Divisions and Federal Electoral Districts (FED), using special estimation techniques. Such estimates, provided they meet certain minimum reliability constraints, are produced and released on a cost-recovery basis. The estimates are used by both federal and provincial governments in allocating funds and other resources among various political and administrative jurisdictions.

The individual sample records from the LFS are available on magnetic tape for interested users. The tapes include demographic information as well as data from the survey questionnaire. In all cases, geographic information is suppressed to ensure the confidentiality of responding households. The micro records are being used increasingly, in conjunction with data from supplementary surveys, in micro-simulation models to assess the impact of changes in government policy. Standard tables and special tabulations can be produced in various machine-readable forms, or in print. Direct access to a wide range of aggregated information is possible through CANSIM, Statistics Canada's machine-readable database and retrieval system.

Standard estimates from the LFS are published monthly in *The Labour Force* (Statistics Canada, Catalogue 71-001). Seasonally adjusted estimates are published for the main characteristics. Requests for special tabulations are filled on a cost-recovery basis.

1.2 An Overview of the Survey

The LFS, due to operational and conceptual problems, excludes from its coverage the Yukon and the Northwest Territories, and residents of Indian reserves and Crown lands. Also excluded are inmates of institutions and full-time members of the Canadian Armed Forces since both groups are considered to exist outside the labour market. These exclusions account for about 2% of the total population of Canada.

The size of the sample remained more or less stable at about 35,000 households per month until the early seventies. During 1976-77, it was increased to about 55,000 households in order to meet the increasing demands for more reliable and detailed data at the provincial level. Due to general restraints on resources, the LFS has also occasionally experienced reductions in its monthly samples. In 1985-86, following two such reductions, the sample comprised about 48,000 households.

The LFS follows a rotating panel sample design, in which households remain in the sample for six consecutive months. The sample is split into six representative sub-samples, and each month one-sixth

pour calculer les indicateurs standard du marché du travail relativement à de petites régions telles les divisions de recensement et les circonscriptions électorales fédérales. Ces estimations, qui doivent répondre à des critères de fiabilité de base, sont produites selon le principe du recouvrement des frais. Le gouvernement fédéral et les gouvernements provinciaux s'en servent pour la répartition des ressources financières et autres entre les diverses compétences politiques et administratives.

Les enregistrements relatifs aux personnes ayant fait partie de l'échantillon de l'enquête sur la population active sont conservés sur bande magnétique et sont disponibles pour les utilisateurs intéressés. Ces bandes contiennent des données démographiques ainsi que les renseignements recueillis au moyen du questionnaire, mais les données géographiques en ont été supprimées afin de protéger l'anonymat des ménages participants. On utilise de plus en plus les micro-enregistrements en même temps que les données tirées des enquêtes supplémentaires pour évaluer, à l'aide de modèles de simulation, l'incidence qu'auraient certains changements s'ils étaient apportés aux politiques gouvernementales. Il est possible de produire sous diverses formes ordinolinguées et sur support papier les tableaux standard ainsi que des totalisations spéciales. Enfin, les personnes intéressées ont accès directement à toute une gamme de données agrégées stockées dans le CANSIM, qui est à la fois la base de données informatique et le système de consultation de Statistique Canada.

Les estimations standard paraissent chaque mois dans la publication *La population active* (Statistique Canada, n° 71-001 au catalogue). Seules les données relatives aux principales caractéristiques sont désaisonnalisées. Des totalisations spéciales sont produites en recouvrement des coûts.

1.2 Aperçu général de l'enquête

Pour des raisons d'ordre opérationnel et théorique, les résidents du Yukon et des Territoires du Nord-Ouest sont actuellement exclus du champ de l'enquête sur la population active, de même que les personnes qui habitent dans des réserves indiennes et sur les terres de la Couronne, les pensionnaires d'un établissement institutionnel et les membres à plein temps des Forces armées canadiennes, qui sont considérés comme n'étant pas sur le marché du travail. Ces personnes représentent environ 2% de la population du Canada.

L'échantillon mensuel, dont la taille s'était maintenue autour de 35,000 ménages jusqu'au début des années 70, a été porté à environ 55,000 ménages en 1976-1977, car la demande de données provinciales plus fiables et plus détaillées se faisait de plus en plus grande. À cause des restrictions qui ont touché l'ensemble des ressources, il a fallu le réduire à l'occasion. Suite aux deux réductions effectuées en 1985-1986, l'échantillon a été réduit de façon à inclure environ 48,000 ménages.

L'EPA est une enquête par panel avec renouvellement de l'échantillon. Celui-ci est divisé en six sous-échantillons représentatifs, dont les ménages font partie pendant six mois consécutifs. Un de ces sous-échantillons est remplacé chaque

of the sample is replaced after it has completed its stay in the sample. This results in a five-sixths month-to-month sample overlap, which makes the design efficient for estimating month-to-month changes. The rotation after six months prevents fatigue and undue respondent burden for selected households.

Data collection for the LFS is carried out during the week following the reference week, which is normally the week containing the fifteenth day of the month. About 780 Statistics Canada interviewers from eight Regional Operations (RO) centres collect the information needed to produce labour force data. This is done through personal interviews in the first month and by telephone interviews in subsequent months. The questionnaires are completed by the interviewer and are sent to the ROs, where data are captured on minicomputers, and then transmitted to Ottawa for processing. The entire LFS operation of data collection, capture, and processing is very tightly monitored and controlled. This enables Statistics Canada to publish the monthly estimates just three weeks after the interview week.

Due to the importance of the statistics produced from the survey and the complexities of the operations involved, various quality evaluation and control programs designed for different phases of the survey are monitored regularly. Furthermore, the sample is partially updated in between the decennial censuses to account for changes in population characteristics and distribution. The sample is also completely redesigned following each decennial census. The latest newly designed sample was introduced in 1985. The earlier sample design and objectives of the redesign, including the main features of the new sample, are outlined briefly below.

1.3 Sample Design Prior to 1985

The salient features of the design used until the redesigned sample was introduced in 1985 are described here to provide a context for the discussion of the design features in the main chapters. Each of Canada's ten provinces was divided into several LFS Economic Regions (ERs), which served as the primary level of stratification. Within each ER the next level of stratification was the designation of areas as Self Representing (SR) and Non-Self Representing (NSR). The SR areas corresponded to urban areas that were large enough to yield an expected sample of 20 dwellings. The rural areas and smaller urban centres made up the NSR areas.

Large cities in the SR areas were further stratified by grouping together contiguous Census Tracts without any regard to optimality of characteristics. Normally, six clusters, corresponding usually to city blocks, were then selected from each stratum using the Rao, Hartley and Cochran (1962) random group method. In addition to this area frame, an open-ended list frame of apartments was maintained in the large cities.

mois, les ménages qui le constituaient ayant fini de participer à l'enquête. Ainsi, les cinq sixièmes de l'échantillon ne changent pas d'un mois à l'autre, ce qui permet de mesurer les variations mensuelles. Par ailleurs, le fait que les ménages sélectionnés sortent de l'échantillon au bout de six mois signifie qu'ils n'ont pas le temps de se lasser d'en faire partie et que le fardeau de réponse n'est pas trop lourd.

La collecte des données de l'EPA est faite pendant la semaine suivant la semaine de référence, qui est normalement celle pendant laquelle tombe le quinzième jour du mois. Environ 780 intervieweurs de Statistique Canada, rattachés à huit bureaux régionaux (BR), recueillent les renseignements nécessaires à la production des données sur la population active. Pour ce faire, ils interviewent chaque ménage de l'échantillon, sur place le premier mois qu'un ménage participe à l'enquête et par téléphone les mois suivants. Les questionnaires remplis par les intervieweurs sont expédiés aux BR où les données sont introduites dans des mini-ordinateurs puis transmises à Ottawa en vue de leur traitement. Tous les aspects des opérations de collecte, de saisie et de traitement des données de l'EPA sont suivis et contrôlés de très près afin que Statistique Canada puisse publier les estimations mensuelles trois semaines à peine après la semaine d'enquête.

Étant donné l'importance des statistiques produites au moyen de cette enquête et la complexité des opérations en cause, chacune des étapes est soumise régulièrement à divers programmes d'évaluation et de contrôle de la qualité. De plus, l'échantillon est partiellement mis à jour entre les recensements décennaux pour tenir compte de l'évolution des caractéristiques et de la distribution de la population et il est entièrement remanié après chacun de ces recensements. L'échantillon prélevé selon le nouveau plan de sondage a été introduit en 1985. Le plan antérieur et les objectifs du remaniement ainsi que les principales caractéristiques du nouvel échantillon sont décrits brièvement ci-après.

1.3 Plan de sondage avant 1985

Les principales caractéristiques du plan de sondage utilisé jusqu'à l'introduction du plan actuel en 1985 sont décrites ici pour fin de comparaisons avec les nouvelles caractéristiques du plan actuel qui seront décrites dans les chapitres suivants. Chacune des dix provinces du Canada était divisée en un certain nombre de régions économiques (RE) de l'EPA, qui constituaient les strates primaires. Le niveau de stratification suivant était la répartition des secteurs qui composaient chaque RE en secteurs autoreprésentatifs (AR) et en secteurs non autoreprésentatifs (NAR). Les secteurs autoreprésentatifs étaient les régions urbaines suffisamment grandes pour que le rendement de l'échantillon y soit de 20 ménages. Les secteurs non autoreprésentatifs étaient formés de régions rurales et de centres urbains plus petits.

Les grandes villes situées dans les secteurs autoreprésentatifs étaient subdivisées, sans souci d'optimalité, en sous-unités constituées de secteurs de recensement contigus. Normalement, six grappes, correspondant en général à des îlots, étaient ensuite choisies dans chaque sous-unité par la méthode des groupes aléatoires de Rao, Hartley et Cochran (1962). Outre cette base areolaire, il existait pour les grandes villes une base constituée de listes ouvertes d'immeubles d'appartements.

The advantages of the selection method used for SR areas were the flexibility it allowed for changes in the sample size (Singh and Drew, 1977) and for updating the sample (Keyfitz, 1951; Platek and Singh, 1977; Drew, Choudhry and Gray, 1978). Sample updating was required in SR areas in order to maintain sampling variances at desired levels by periodically adjusting the sample to reflect uneven population growth.

In the NSR areas, stratification based on industry classification was carried out within each Economic Region, subject to the condition that strata should be contiguous land areas. Within strata, Primary Sampling Units (PSUs) were delineated such that each PSU represented its stratum with respect to the rural to urban population ratio and important labour force characteristics. Although the rural portions of the PSU were made contiguous by grouping neighbouring rural Enumeration Areas, contiguity of urban portions could not always be attained due to the restriction placed on maintaining the rural to urban population ratio. Frequently, larger urban centres had to be shared among several PSUs within a stratum.

Two or more PSUs were selected from each stratum using the randomized probability proportional to size (PPS) systematic sampling method (Hartley and Rao, 1962). Within each selected PSU, the urban and rural portions were sampled independently. In the urban portion of the selected PSU, a two stage sample of clusters and dwellings was selected. In the rural portion a three stage design was used: secondary sampling units (consisting of Census Enumeration Areas or combinations), followed by clusters (identifiable land areas with up to 20 dwellings) and then dwellings as the final sampling units. Except for the last stage, randomized PPS systematic sampling was used in the selections. The last stage units, namely dwellings, were selected systematically in both SR and NSR areas.

Following a personal interview in the first month in the sample, telephone interviewing was the normal method of data collection for households in their second to sixth month in SR areas. In the NSR areas, however, all interviewing was carried out in person due to concerns about public acceptance of telephone interviewing in rural areas, particularly in the light of party lines.

1.4 Objectives of the Redesign

Following each decennial population census, the Canadian Labour Force Survey has undergone a sample redesign to reflect changes in population characteristics and to respond to changes in the information needs to be satisfied by the survey. The current redesign program culminated with the introduction of a new sample at the beginning of 1985. The program included extensive consultation involving the

L'avantage que présentait la méthode de sélection utilisée pour les secteurs AR était sa souplesse lorsqu'il fallait modifier la taille de l'échantillon (Singh et Drew 1977) et mettre celui-ci à jour (Keyfitz 1951, Platek et Singh 1977 et Drew, Choudhry et Gray 1978). Il fallait tenir l'échantillon à jour dans les secteurs AR pour maintenir les variances d'échantillonnage aux niveaux voulus, ce qui était fait en rajustant périodiquement l'échantillon pour tenir compte de la croissance inégale de la population.

Dans les secteurs NAR, la stratification était faite à l'intérieur de chaque région économique en fonction de la répartition des activités économiques, mais il fallait que les strates soient des superficies contiguës. Dans chaque strate, les unités primaires d'échantillonnage (UPE) devaient être délimitées de telle manière que chaque UPE soit représentative de sa strate du point de vue du ratio de la population rurale à la population urbaine et de certaines caractéristiques importantes de la population active. Il était possible de faire en sorte que les parties rurales des UPE soient contiguës en groupant des secteurs de dénombrement voisins, mais ce but ne pouvait pas toujours être atteint avec les centres urbains, vu la nécessité de maintenir le ratio de la population rurale à la population urbaine. On a souvent dû répartir les grands centres urbains entre plusieurs UPE d'une même strate.

Deux UPE ou plus étaient choisies dans chaque strate par la méthode d'échantillonnage systématique avec classement aléatoire et probabilité proportionnelle à la taille (Hartley et Rao 1962). Les parties urbaines et rurales de chaque UPE choisie étaient échantillonnées séparément. Dans la partie urbaine des UPE, on s'est servi d'un plan d'échantillonnage à deux degrés pour choisir les grappes et les logements. Dans la partie rurale, on a appliqué un plan d'échantillonnage à trois degrés pour les unités secondaires, c'est à dire pour prélever les secteurs de dénombrement du recensement ou groupes de tels secteurs, les grappes (en l'occurrence, des superficies habitées par un maximum de 20 ménages) et les logements (dernières unités d'échantillonnage). Exception faite de la dernière étape, l'échantillon a été prélevé par la méthode d'échantillonnage systématique avec classement aléatoire et probabilité proportionnelle à la taille. Tant dans les secteurs AR que NAR, on a eu recours à l'échantillonnage systématique pour les dernières unités, soit les logements.

Pour la collecte des données, la méthode normale consistait à faire des interviews téléphoniques auprès des ménages dans les secteurs AR du deuxième au sixième mois où ceux-ci participaient à l'enquête, tandis que dans les secteurs NAR toutes les interviews étaient faites sur place, la population des régions rurales étant généralement moins favorable aux interviews téléphoniques, surtout compte tenu du fait que dans beaucoup d'endroits les lignes téléphoniques sont partagées par plusieurs abonnés.

1.4 Objectifs du remaniement

Après chaque recensement décennal de la population, on remanie le plan de sondage de l'enquête sur la population active du Canada pour tenir compte de l'évolution des caractéristiques de la population et des besoins en information que cette enquête cherche à satisfaire. Le dernier programme de remaniement a abouti à l'introduction d'un nouvel échantillon au début de l'année 1985. Dans le cadre de ce programme, on a procédé à des consultations poussées visant

reassessment not only of the survey's principal role as a provider of current labour market information, but also of its use as a central vehicle for conducting household surveys within Statistics Canada. The following three objectives were established for the sample redesign:

(i) Improved subprovincial labour market data.

This objective was established in consultation with the statistical focal points within each of Canada's ten provinces, and with key federal user departments, through annual federal/provincial conferences on labour statistics and bilateral follow-ups. In general, these consultations revealed satisfaction with data reliability for provincial and national data, as well as a strong desire for improved subprovincial data. Specific data reliability requirements adopted for the redesigned sample were as follows:

- (a) for Canada and each of the ten provinces: no reduction in current reliability for monthly estimates of level and estimates of month-to-month change in total employment and unemployment.
- (b) for the 24 Census Metropolitan Areas as defined by the 1981 Census: monthly estimates of unemployed with coefficients of variations (CVs) of 20% or less.
- (c) for 66 subprovincial Economic Regions defined in consultation with the provinces: monthly estimates of unemployed with CVs of 25% or less.
- (d) for cities with a population of 60,000 or more in Quebec and Ontario, and 25,000 or more in the remaining provinces: quarterly estimates of unemployed with CVs of 25% or less.

(ii) Enhance the LFS as a general vehicle for the conduct of household surveys.

During the revision of the LFS in the 1970s, changes were made to the survey operations and systems that expanded the capacity for other household surveys to use the existing LFS infrastructure. In the ensuing decade both the number and diversity of such surveys increased substantially. This objective related to explicitly taking account of such general applications in the survey design, for example, by making the stratification less labour force specific.

(iii) Improve the cost-efficiency of the survey design.

From the cost-efficiency perspective, the potential was recognized for major savings if the telephone interviewing procedure followed in SR areas could be extended to NSR areas. In parallel with the investigation of more telephoning, it was decided that sample design parameters (such as the number of stages of sampling and the allocation of the sample to stages) should be re-optimized, particularly in NSR areas.

à réévaluer non seulement le contrôle principal de l'enquête, c'est-à-dire de permettre la production de données actuelles sur le marché du travail, mais aussi l'utilisation qui en est faite à Statistique Canada comme principal moyen pour effectuer les enquêtes-ménages. Le remaniement du plan de sondage avait les trois grands objectifs suivants :

(i) Améliorer les données infraprovinciales sur le marché du travail

Cet objectif a été établi à la suite de consultations auprès des points de contact dans les dix provinces canadiennes et des principaux ministères et organismes fédéraux utilisateurs, dans le cadre des Conférences fédérales-provinciales sur la statistique du travail et d'échanges bilatéraux subséquents. Il est ressorti de ces consultations que les personnes intéressées étaient généralement satisfaites des données provinciales et nationales sur le plan de la fiabilité, mais qu'elles souhaitaient fortement que les données infraprovinciales soient améliorées. Le nouveau plan de sondage devait donc remplir les conditions suivantes à cet égard :

- a) Pour le Canada et les dix provinces, permettre de maintenir le degré de fiabilité des estimations mensuelles du niveau de l'emploi et du chômage et des estimations de la variation mensuelle de ce niveau.
- b) Pour les 24 régions métropolitaines de recensement telles que définies pour le recensement de 1981, permettre de produire des estimations mensuelles du chômage dont le coefficient de variation (CV) ne dépasserait pas 20%.
- c) Pour 66 régions économiques à l'intérieur des provinces choisies avec l'accord de ces dernières, permettre de produire des estimations mensuelles du chômage dont le CV ne dépasserait pas 25%.
- d) Pour les villes de 60,000 habitants ou plus du Québec et de l'Ontario et de 25,000 habitants ou plus des autres provinces, permettre de produire des estimations trimestrielles du chômage dont le CV ne dépasserait pas 25%.

(ii) Faire de l'EPA un meilleur outil pour la réalisation d'enquêtes-ménages.

Lors de la révision de l'EPA dans les années 70, on a apporté des modifications aux opérations d'enquête et aux systèmes informatiques afin de permettre à un plus grand nombre d'enquêtes-ménages d'utiliser l'infrastructure en place. Le nombre et la variété de ces enquêtes se sont accrues de façon considérable pendant les dix années qui ont suivi. Cet objectif signifiait qu'on tiendrait compte, au moment du remaniement du plan de sondage, des applications générales qui en seraient faites, en faisant en sorte, par exemple, que la stratification n'ait pas comme unique critère les caractéristiques de la population active.

(iii) Rendre le plan de sondage plus économique.

Il a été reconnu que de grandes économies pourraient être réalisées si la méthode de collecte des données par interview téléphonique utilisée dans les secteurs AR était appliquée dans les secteurs NAR. Parallèlement à l'étude de cette question, il a été décidé qu'il faudrait optimiser de nouveau les paramètres du plan de sondage comme le nombre de degrés d'échantillonnage et la répartition de l'échantillon selon le degré, particulièrement dans les secteurs NAR.

1.5 Scope of the Report

This report is intended to be a reference source on the methodological and operational aspects of the LFS and is aimed at program planners, large scale survey designers and managers, analytic users, researchers and consultants. The survey design and methodology are discussed in detail with supplementary references indicated where appropriate. To complement this description of the survey mechanism a separate report entitled 'Guide to Labour Force Survey Data', with emphasis on subject matter aspects and the data is available from the LFS has been prepared. The sample design and various operational procedures described in different chapters of this report are as of January 1985, except for Estimation Methodology in Chapter 8 and developmental programs discussed in the last chapter. In these cases the report includes changes stemming from ongoing research activities in search of more cost effective data collection methods and better estimation techniques that were incorporated as late as 1988-89, i.e. until the finalization of this report.

There are ten additional chapters in this report. Chapter 2 describes various levels of stratification, choice of stratification variables and corresponding algorithms, and the stratification system adopted in the Self Representing and Non-Self Representing areas of an Economic Region.

Chapters 3 and 4 discuss in detail the design considerations in the Self Representing and Non-Self Representing areas respectively including the choice and formation of sampling units at different stages of selection, the sampling methods adopted, and the methods of updating the sample.

Rotation of the sample occurs at each stage of the multistage design used in the LFS. Chapter 5 describes the method used in the allocation of the overall sample to the various provinces and subprovincial areas. Chapter 6 explains how this rotation (replacement of units) is accomplished in an unbiased manner, taking into account the cost of rotation and the response burden.

Chapter 7 identifies the role of various divisions involved in the interdisciplinary environment governing the overall operation of the LFS. It also presents an overall picture of the operations involved in a typical survey month by describing the operational aspects of sampling, data collection, and major steps in data processing and dissemination.

Chapter 8 describes the new estimation and variance estimation methodology introduced in 1988-89 as a result of research work undertaken after the introduction of the new sample in 1985. The new estimation method, which uses a regression approach, was found to be at least as good as the raking ratio

1.5 Objet de la publication

Cette publication se veut un ouvrage de référence sur la méthodologie et les opérations de l'EPA et elle s'adresse aux planificateurs de programmes, aux personnes qui conçoivent de grandes enquêtes ainsi qu'à celles chargées de leur réalisation, aux utilisateurs des données à des fins d'analyse, aux chercheurs et aux conseillers. Le plan et la méthodologie de l'enquête sont traités en détail et le lecteur est renvoyé à des documents de référence supplémentaires au besoin. Un document distinct intitulé "Guide des données de l'enquête sur la population active" complète la description du fonctionnement de l'enquête présentée ici; l'accent y est mis sur les concepts et définitions et sur les données produites dans le cadre de l'EPA. Le plan de sondage et les procédures opérationnelles décrites dans les différents chapitres datent de janvier 1985 à l'exception de la méthodologie d'estimation présentée au chapitre 7 et des programmes de développement dont il est question au dernier chapitre. Il s'agit là de modifications découlant des travaux de recherche qui se font continuellement pour trouver des méthodes de collecte de données plus économiques et des techniques d'estimation plus efficaces. Certaines de ces modifications ont été apportées aussi tard qu'en 1988-1989, alors qu'on mettait la dernière main au manuscrit.

Cette publication contient onze chapitres. Le chapitre 2 décrit les divers niveaux de stratification, le choix des variables de stratification et les algorithmes qui s'y rapportent ainsi que le système de stratification adopté pour les secteurs auto-représentatifs et non auto-représentatifs d'une région économique.

Les chapitres 3 et 4 traitent en détail des facteurs qui sont intervenus dans l'élaboration du plan de sondage en ce qui a trait aux secteurs auto-représentatifs et non auto-représentatifs respectivement, notamment le choix et la formation des unités aux différents degrés d'échantillonnage ainsi que les méthodes d'échantillonnage et de mise à jour de l'échantillon adoptées.

Le renouvellement de l'échantillon est effectué à chaque étape du processus d'échantillonnage à plusieurs degrés de l'EPA. Au chapitre 5, nous décrivons la méthode par laquelle l'échantillon global est réparti entre les provinces et régions infraprovinciales et, au chapitre 6, nous expliquons comment le renouvellement (le remplacement des unités) est réalisé sans risque de biais et en tenant compte du coût de cette opération et du fardeau de réponse.

Au chapitre 7, nous indiquons le rôle joué par les diverses divisions qui constituent le contexte interdisciplinaire dans lequel les opérations de l'EPA sont accomplies. Nous donnons également une vue d'ensemble des opérations effectuées pendant un mois d'enquête typique, c'est-à-dire que nous décrivons les aspects opérationnels de l'échantillonnage, de la collecte des données et des principales étapes du traitement et de la diffusion des données.

Au chapitre 8, nous décrivons les nouvelles méthodes d'estimation et d'estimation de la variance introduites en 1988-1989 à la suite des travaux de recherches entrepris après l'adoption du nouveau plan en 1985. On a trouvé que, en ce qui concerne les caractéristiques de la population active, la nouvelle méthode, fondée sur l'analyse de la régression, est

method for the Labour Force (LF) characteristics and at the same time achieved the desired consistency among individual and family related statistics. The raking ratio method of estimation and the corresponding variance estimation technique used between 1984-85 and 1988-89 are also briefly described.

Important quality measures and quality monitoring and evaluation programs for the LFS are discussed in Chapter 9. Chapter 10 briefly describes the use of the LFS as a central vehicle for general application to other household surveys.

Chapter 11 describes some of the research initiatives underway at the time this report was prepared in order to find ways of responding to changing demands for data and to take advantage of new technology in data collection and processing. Several appendices, each dealing with a special topic, and other survey materials, are included at the end.

au moins aussi bonne que la méthode itérative du quotient et qu'elle permet en même temps d'obtenir le degré de cohérence voulu entre les statistiques ayant trait aux individus et celles se rapportant à la famille. La méthode du quotient et la technique d'estimation de la variance correspondante utilisées entre 1984-1985 et 1988-1989 sont décrites brièvement.

Les mesures et les programmes importants de vérification et d'évaluation de la qualité appliqués à l'EPA sont décrits au chapitre 9. Au chapitre 10, nous parlons brièvement de la manière dont on se sert de l'EPA pour réaliser d'autres enquêtes-ménages.

Le chapitre 11 porte sur les travaux de recherche en cours au moment de la rédaction du présent ouvrage. Ceux-ci visent à trouver des moyens de réagir face à l'évolution de la demande de données et d'appliquer les nouvelles technologies à la collecte et au traitement des données. Les annexes, dont chacune traite d'un sujet particulier, et les autres documents d'enquête sont inclus à la fin de la publication.

Chapter 2

Stratification

2.1 Introduction

In the design of sample surveys, stratification refers to the process of partitioning the sampling frame into subsets called strata, each of which are sampled separately. Stratification is usually carried out in such a way as to reduce the variability of the estimates of population characteristics. This is done by grouping together units which are homogenous with respect to the population characteristics of interest. Another advantage of stratification is that, depending upon the information available, sampling and estimation procedures may differ from stratum to stratum. Further, in a continuous survey like the Labour Force Survey, stratification provides the added flexibility of updating or redesigning the sample in a specified stratum or group of strata, without affecting the design in the remaining strata.

Each of the ten provinces of Canada is divided into a number of LFS Economic Regions. Economic Regions were originally established by the Department of National Defense in the 1950s as subprovincial areas of similar economic structure. Since the LFS design introduced after the 1961 Census, Economic Regions have been used as primary strata. Initially this was done to ensure better geographical representation of the sample within provinces, and estimates were published only for provinces or groups of provinces. But, since the increase in the size of the LFS sample during the late 1970s, estimates have been published for individual or groups of Economic Regions. Since their availability, estimates at the Economic Region level have been used extensively by the provinces and by federal government departments; for example, Employment and Immigration Canada, uses them in the administration of the Unemployment Insurance Act.

As a result of the increased use of estimates at the Economic Region level, a good deal of consultation took place during the initial stages of the current sample redesign, with the provinces and with federal government departments regarding the choice of Economic Region boundaries for the sample. These discussions led to the adoption of completely new Economic Region boundaries for Ontario, Quebec and Alberta; there were little or no changes in the other provinces. Maps and descriptions of Economic Region boundaries for each province are given in Appendix E.

Within Economic Regions, the next level of stratification is the designation of areas as **Self Representing** (SR) or **Non-Self Representing** (NSR).

Chapitre 2

Stratification

2.1 Introduction

Dans le plan des enquêtes par sondage, la stratification est le processus qui consiste à diviser la base de sondage en sous-ensembles appelés des strates, dont chacune est échantillonnée séparément. On procède habituellement à la stratification de manière à réduire la variabilité des estimations relatives aux caractéristiques de la population, c'est-à-dire en groupant dans une même strate les unités qui sont homogènes du point de vue des caractéristiques-clés. La stratification présente d'autres avantages, dont celui de permettre de varier la méthode d'échantillonnage et d'estimation d'une strate à l'autre en fonction des renseignements disponibles. En outre, pour une enquête permanente comme l'EPA, la stratification permet de mettre l'échantillon à jour ou de remanier le plan de sondage d'une strate donnée ou de plusieurs groupes de strates sans toucher au plan de sondage des autres strates.

Chacune des dix provinces canadiennes est divisée en un certain nombre de régions économiques de l'EPA. À l'origine, lorsque les régions économiques avaient été établies par le ministère de la Défense nationale dans les années 50, il s'agissait de secteurs infraprovinciaux de structure économique semblable. Au remaniement du plan de sondage de l'EPA qui a suivi le recensement de 1961, les régions économiques sont devenues les strates primaires. Ce changement a été apporté au départ pour assurer une meilleure représentation géographique de l'échantillon dans les provinces, les estimations publiées à l'époque ne se rapportant qu'aux provinces et groupes de provinces. Mais depuis que la taille de l'échantillon a été agrandie à la fin des années 70, il a été possible de publier des estimations au niveau des régions économiques individuelles ou regroupées. Maintenant que ces données existent, les provinces et les ministères et organismes fédéraux s'en servent énormément; Emploi et Immigration Canada, par exemple, s'en sert pour appliquer la Loi sur l'assurance-chômage.

Compte tenu de cette forte utilisation des estimations relatives aux régions économiques par les provinces et les ministères et organismes fédéraux, ceux-ci ont été amplement consultés pendant les étapes initiales du remaniement du plan de sondage au sujet des limites à adopter pour les régions économiques. Ces consultations consultations ont abouti à un découpage totalement nouveau pour le Québec, l'Ontario et l'Alberta, mais à peu ou pas de changement dans le reste du pays. L'annexe C contient des cartes, accompagnées d'explications, qui illustrent les régions économiques constituées dans chaque province.

Le niveau de stratification suivant a consisté à déterminer quels secteurs dans les régions économiques étaient **autoreprésentatifs** et lesquels étaient **non autoreprésentatifs**.

The SR areas correspond to urban areas sufficiently large to yield a sample of at least fifty dwellings, the minimum felt acceptable for an interviewer's assignment. Due to the importance placed on improved reliability of data for smaller Economic Regions in the redesign of the survey, sampling rates, and hence the minimum sizes for urban areas to be classified as SR, differed from one Economic Region to the next.

The rural areas and smaller urban centres not qualifying as SR comprise the NSR areas. Within both the NSR and SR portions of Economic Regions, a type of optimal stratification was carried out. Because the LFS is used not only to provide information on labour force characteristics but also as a general vehicle for various other household surveys, one of the principal objectives of the redesign was to increase the flexibility of the LFS for general applications. Stratification was considered a means of improving efficiency for general applications (as well as for variables of particular interest to the LFS) through the application of more rigorous procedures than those used in the old design.

Drawing upon findings of studies carried out at the United States Bureau of the Census in the redesign of its Current Population Survey, Statistics Canada decided to use a non-hierarchical, multivariate clustering algorithm for stratification, originally proposed by Friedman and Rubin (1967). The basic algorithm was modified to permit formation of compact and contiguous strata, and also to permit delineation of heterogeneous but geographically compact *Primary Sampling Units* (PSUs) within NSR strata.

In addition to the delineation of more statistically efficient strata and NSR Primary Sampling Units, the automated system resulted in considerable savings in both the human resources and time required to complete these steps. PSU delineation, in particular, had been a time-consuming step when done manually during earlier redesigns.

Section 2 discusses the stratification algorithm and considerations leading to the choice of stratification variables. Section 2.3 describes the stratification adopted in SR areas. For NSR areas, both the stratification and delineation of PSUs are discussed together in section 2.4 due to the similarity of the two operations. More details on the stratification algorithm and studies to evaluate the stratification adopted for the redesigned sample can be found in Drew, Belanger and Foy (1985).

Les secteurs autoreprésentatifs (AR) sont des régions urbaines suffisamment grandes pour que le rendement de l'échantillon y soit d'au moins cinquante ménages, ce qui constitue le nombre minimum de ménages à confier à un intervieweur. Comme on tenait beaucoup à profiter du remaniement du plan de sondage pour augmenter la fiabilité des données sur les petites régions économiques, on a varié d'une région économique à une autre les fractions de sondage et par conséquent les seuils de taille minimum pour qu'une région urbaine soit considérée comme un secteur autoreprésentatif.

Les secteurs non autoreprésentatifs (NAR) sont composés des régions rurales et des centres urbains trop petits pour être comptés parmi les secteurs autoreprésentatifs. Tant dans les secteurs AR que dans les secteurs NAR des régions économiques, on a procédé à une stratification optimale. Comme on ne se sert pas seulement de l'EPA pour produire des données sur les caractéristiques de la population active, mais aussi pour réaliser d'autres enquêtes-ménages, un des principaux objectifs du remaniement du plan de sondage était de faire en sorte que l'EPA puisse se prêter plus facilement à des applications générales. On a pensé qu'en appliquant des méthodes plus rigoureuses que dans l'ancien plan on pourrait se servir de la stratification pour augmenter l'efficacité de l'échantillonnage à la fois pour les applications générales et pour les variables qui intéressent particulièrement l'EPA.

En se basant sur les résultats d'études effectuées par le United States Bureau of the Census en vue de la refonte de la Current Population Survey (enquête sur la population actuelle), on a décidé d'utiliser un algorithme de classification non hiérarchique et multidimensionnel proposé à l'origine par Friedman et Rubin (1967) à des fins de stratification. L'algorithme de base a été modifié afin de permettre la formation de strates compactes et contiguës et aussi d'unités primaires d'échantillonnage (UPE) hétérogènes mais géographiquement compactes dans les strates non autoreprésentatives.

Non seulement le système automatisé permet-il de créer des strates et des UPE NAR plus efficaces du point de vue statistique, mais en plus il a entraîné des économies considérables sur le plan des ressources humaines et du temps requis pour accomplir cette tâche. La délimitation des UPE en particulier prenait beaucoup de temps lorsqu'elle était faite à la main, selon les plans de sondage antérieurs.

A la section 2, il sera question de l'algorithme de stratification et des raisons qui ont motivé le choix des variables de stratification. La section 3 décrit la stratification adoptée dans les secteurs AR. Pour ce qui est des secteurs NAR, la stratification et la délimitation des UPE sont toutes deux traitées à la section 4 du fait que les deux opérations se ressemblent. Le lecteur trouvera plus de renseignements sur les algorithmes de stratification et sur les études d'évaluation de la stratification adoptée dans le nouveau plan de sondage dans l'article de Drew, Bélanger et Foy (1985).

2.2 Stratification Algorithm

The algorithm is designed to partition the units into groups (strata) that are as homogeneous as possible for several variables of interest, by minimizing the weighted sums of the squares for sampling with PPS within each group.

The objective function to be minimized is SW , the within-group sum of squares (SW_i) weighted for all p variables, $i = 1, 2, \dots, p$, that is,

$$SW = \sum_{i=1}^p w_i SW_i$$

The weighted between-group sum of squares (SB) and the weighted total sum of squares (ST) are defined similarly, such that

$$ST = SW + SB$$

and

$$ST_i = SW_i + SB_i.$$

Note that SW_i , SB_i and ST_i are defined for the sampling with probability proportional to size. Further, the stratification index for variable i , is defined as

$$I_i = \frac{SB_i}{ST_i}, i = 1, 2, \dots, p.$$

A high index value indicates a good stratification (clustering).

One way of identifying the best clustering would be to generate all the possible partitions of N -units into L -groups and then simply select the one that minimizes the objective function SW . This approach is rarely feasible because the number of possible partitions may be unmanageably large.

The Friedman and Rubin (1967) algorithm begins with any partition of the N -units into L -groups. In an iteration of the algorithm, each unit is examined in turn and assigned to the group (including its current group), which results in the lowest value for the objective function. After several iterations, the algorithm reaches a point at which no move of a single unit will produce a reduction in the objective function. This point is described as a local minimum of the objective function because it is dependent on the starting partition. Another starting partition might have achieved an even lower value of the objective function. Multiple random starting points (30 per stratum) were used to move beyond local minima.

Previous LFS sample designs used strata composed of contiguous geographic units; that is, each unit in a given stratum had to be touching at least one other unit in the same stratum. One reason was the assumption that such strata would retain the efficiency of the sample design for a longer period of time than would strata made up of non-contiguous units.

2.2 Algorithme de stratification

L'algorithme a pour objet de partitionner les unités en groupes (strates) aussi homogènes que possible du point de vue d'un certain nombre de variables-clés en minimisant les sommes des carrés pondérées pour l'échantillonnage avec PPT dans chaque groupe.

La fonction objective devant être minimisée est SW , la somme des carrés intraclasse (SW_i) pondérée pour toutes les variables p , $i = 1, 2, \dots, p$, c'est-à-dire

$$SW = \sum_{i=1}^p w_i SW_i$$

La somme des carrés interclasse pondérée (SB) et la somme totale des carrés pondérée (ST) sont définies de façon semblable, de telle sorte que

$$ST = SW + SB$$

et

$$ST_i = SW_i + SB_i.$$

Noter que SW_i , SB_i et ST_i sont définies pour l'échantillonnage avec probabilité proportionnelle à la taille. De plus, l'indice de stratification pour la variable i est défini

$$I_i = \frac{SB_i}{ST_i}, i = 1, 2, \dots, p.$$

Un indice élevé dénote une bonne stratification (homogénéité).

Pour obtenir la meilleure classification automatique, on pourrait générer toutes les partitions possibles de N -unités en L -groupes et simplement choisir celle qui minimise la fonction objectif SW . Ceci est rarement réalisable parce que le nombre de partitions possibles peut atteindre un ordre de grandeur excessif.

L'algorithme de Friedman et Rubin (1967) est un algorithme heuristique qui commence par une répartition de N -unités L -groupes, choisis au hasard. À chaque itération de l'algorithme, chaque unité est examinée à tour de rôle et attribuée au groupe (y compris le groupe dont elle fait partie) qui permet d'obtenir la fonction objective à valeur minimum. Après plusieurs itérations, on atteint un point où aucun déplacement d'unité ne réduit la fonction objective. Ce point est décrit comme étant le minimum local de la fonction objective parce qu'il dépend de la partition initiale. Un autre mode de répartition initiale aurait peut-être permis d'atteindre une fonction objective de valeur encore plus faible. On a pris plusieurs points de départ aléatoires (30 par strates) pour trouver un des minimum locaux les plus faibles.

Dans les plans de sondage antérieurs de l'EPA, les strates étaient composées d'unités géographiques contiguës, c'est-à-dire que, dans une strate donnée, chaque unité devait toucher au moins une autre unité. Cette condition avait été retenue en partie parce qu'on supposait que l'efficacité du plan de sondage serait assurée pendant plus longtemps avec de telles strates qu'avec des strates composées d'unités non contiguës.

To assess this assumption and adopt the best possible stratification, two means were considered for taking geography into account in the stratification. The first method consists of the use of geographic variables (centroids) as variables of interest. This method uses two centroids, which are transformations of longitude and latitude. It yields compact strata, that is, strata in which the distance between units is made minimal by minimizing the usual within-group sum of squares of the centroids. However, the minimization is tempered by minimization of the other variables of interest. Moreover, there is no assurance that these strata will be composed of contiguous units.

A second approach was developed making use of contiguity vectors, which for each unit contained an identification of all units contiguous to it. This approach guaranteed contiguous, but not necessarily compact, strata. Under this approach optimization is performed as described above, but beginning, in this case, with a random starting partition that is contiguous. The generation of such initial partitions is done automatically by the stratification system. The movement of units between strata is permitted only if, in addition to reducing the objective function, the contiguity of both strata is preserved.

It was decided to standardize the stratification variables by introducing weighting factors inversely proportional to the total sum of squares of each variable. This standardization makes it possible to obtain a comparable contribution by each variable to the stratification. Flexibility was retained after standardization to assign one or more variables of greater importance in relation to the other variables in the optimization by specifying a weight greater than one. For example, a variable weight of two would assign double importance. This feature was used to assign sufficient importance to centroids, without unduly affecting the optimization for non-geographic variables. With experimentation it was found that generally the combined use of contiguity vectors and centroids with a weight of 3 relative to other variables offered a good compromise. To assist in manually reviewing and assessing the geographic suitability of stratification, the system featured a plotting of centroids of the stratification units comprising each stratum.

The stratification algorithm that was adopted did not place any limit on the number of stratification variables. Stratification variables were chosen based on results from a study of 11 Economic Regions across Canada, in which strata were delineated using 1971 Census data, and assessed using 1981 Census data. It was found that a general stratification using 16 census variables resulted in the best stratification for estimation of labour force characteristics and for other characteristics as well. The 16 variables included the following: income; secondary education; eight variables related to employment (total persons employed, and persons employed by seven major industry groupings of agriculture, forestry and fisheries, mines, manufacturing, constructions, transportation and services); four variables related to housing (number of owned

Pour vérifier cette hypothèse et pouvoir adopter la meilleure stratification possible, on a envisagé deux façons de tenir compte du critère géographique au moment de la stratification. La première méthode consiste à prendre les centroïdes comme variables-clés. Il s'agit donc d'utiliser deux variables géographiques (les centroïdes), qui sont une variante de la longitude et de la latitude. Cette méthode donne des strates compactes, c'est-à-dire des strates à l'intérieur desquelles la distance entre les unités est réduite au maximum en minimisant la somme des carrés intraclasse des centroïdes. Cette minimisation est cependant tempérée par celle des autres variables-clés, et il n'y a aucune garantie que les strates ainsi obtenues seront composées d'unités contiguës.

La seconde méthode élaborée repose sur l'utilisation de vecteurs de contiguïté, le vecteur relatif à chaque unité identifiant toutes les unités qui y sont contiguës. Cette méthode garantit des strates contiguës, mais pas nécessairement compactes. On obtient l'optimisation de la façon décrite plus haut, mais en commençant cette fois-ci par une partition initiale aléatoire d'unités contiguës. Le système de stratification génère automatiquement de telles partitions initiales. Le mouvement des unités entre les strates est permis uniquement si, en plus de réduire la fonction objective, il permet de préserver la contiguïté des deux strates en cause.

Il a été décidé de normaliser les variables de stratification en introduisant des facteurs de pondération inversement proportionnels à la somme totale des carrés de chaque variable. De cette manière, chaque variable jouait un rôle comparable dans la stratification. Il était toujours possible, après la normalisation, de donner à une ou plusieurs variables davantage d'importance par rapport aux autres en leur attribuant un poids supérieur à 1 à l'étape de l'optimisation. Ainsi, on doublerait l'importance d'une variable en lui donnant un poids égal à 2. On s'est servi de ce moyen pour donner suffisamment d'importance aux variables géographiques (les centroïdes) sans pour autant trop nuire à l'optimisation pour les variables non géographiques. On a trouvé après un certain nombre d'essais que l'utilisation conjuguée des vecteurs de contiguïté et des centroïdes auxquels on avait attribué un poids de 3 par rapport aux autres variables constituait en général un bon compromis. Le système comportait une fonction de représentation graphique des centroïdes des unités de stratification comprises dans chaque strate, ce qui permettait de revoir la stratification et de déterminer si elle convenait du point de vue géographique.

L'algorithme de stratification adopté n'imposait aucune limite quant au nombre de variables de stratification. Celles-ci ont été choisies d'après les résultats d'une étude portant sur 11 régions économiques prises dans l'ensemble du Canada et dont les strates avaient été délimitées en fonction des données du recensement de 1971 et vérifiées au moyen des données du recensement de 1981. On a trouvé que la meilleure stratification pour l'estimation des caractéristiques de la population active et d'autres caractéristiques également était une stratification générale faisant entrer en jeu 16 variables de recensement. Ces variables étaient : le revenu; les études secondaires; 8 variables ayant trait à l'emploi (nombre total de personnes occupées et nombre de personnes occupées selon sept grands groupes d'activités économiques soit l'agriculture, l'exploitation forestière et la pêche, les mines, le secteur manufacturier, le bâtiment, le transport et les

dwelling, total gross rent, number of one person households, and number of two person households; and two geographic variables (centroids).

2.3 Stratification in Self Representing Areas

Individual urban areas designated as Self Representing are known as Self Representing Units (SRUs). As the name implies, the SRU is the primary level of stratification of the SR sample. The extent of further stratification is largely dependent on the size of individual SRUs.

Within the larger SRUs, the first level of stratification is due to the establishment of two separate sampling frames: a list frame of high rise apartment buildings, and an area frame covering remaining dwelling units. The smaller SRUs feature only an area frame. The further stratification within the area frame and apartment frame are described separately below.

(a) Stratification of the Area Frame

In the Census Metropolitan Areas (CMA), Census Tracts were adopted as stratification units. These are geostatistical units with populations between 3,000 and 5,000 persons, which are stable from one census to the next. Both their size and their stability make them practical units for stratification. Primary and secondary strata were formed using the automated stratification system described in the previous section.

The strata were made contiguous and compact by using contiguity vectors and weights of three on the geographic variables relative to other variables. Primary strata were made large enough to yield a sample take that was suitable for two interviewers. Secondary strata (4 to 5 per primary), were formed without regard to geographic constraints.

The two levels of stratification in the larger SRUs had appeal on both operational and technical grounds. The relaxing of geographic constraints over those existing in the old design permitted greater optimality, while the retention of contiguity at a higher level provided a suitable unit for sample updating later in the decade, and facilitated the planning of interviewer assignments. Also, in the old design, SR strata were usually covered entirely by a single interviewer, and hence the variance estimates did not reflect the correlated response variance component of total variance. To the extent that within strata interviewer assignments are geographic and secondary strata are non-geographic, an interpenetration of strata and interviewer assignments has been achieved in the new design without incurring any additional data collection costs, resulting in this component being better reflected in the variance estimates.

services); 4 variables ayant trait au logement (nombre de logements possédés, loyer brut total, nombre de ménages à une personne et nombre de ménages à deux personnes); et 2 variables géographiques (les centroides).

2.3 Stratification dans les secteurs autoreprésentatifs

On appelle unités autoreprésentatives (UAR) les régions urbaines classées dans la catégorie des secteurs autoreprésentatifs. Comme leur nom l'indique, les UAR constituent le niveau primaire de stratification de l'échantillon autoreprésentatif. Le nombre de niveaux de stratification de plus auxquels on procède dépend de la taille de l'UAR.

Dans les grandes UAR, le premier niveau de stratification se justifie par la nécessité d'établir deux bases de sondage distinctes : une liste d'immeubles d'appartements et une base aréolaire comprenant les logements restants. Dans les petites UAR, il n'y a qu'une base aréolaire. La stratification des bases aréolaires et des bases d'immeubles d'appartements est décrite ci-après.

(a) Stratification de la base aréolaire

Dans les régions métropolitaines de recensement, on a pris comme unités de stratification les secteurs de recensement. Il s'agit d'unités géostatistiques dont la population varie entre 3,000 et 5,000 habitants et qui ne changent pas d'un recensement à l'autre. Leur taille et leur stabilité en font des unités pratiques du point de vue de la stratification. Les strates primaires et secondaires ont été formées au moyen du système de stratification automatisé décrit dans la section précédente.

On a créé des strates contiguës et compactes en utilisant les vecteurs de contiguïté et en donnant aux variables géographiques un poids de 3 par rapport aux autres variables. On a formé des strates primaires assez grandes pour que le rendement de l'échantillon corresponde à la tâche de deux intervieweurs. Les strates secondaires (4-5 par strate primaire) ont été constituées sans contraintes géographiques.

Le fait de procéder à deux niveaux de stratification dans les grandes UAR présentait des avantages tant sur le plan opérationnel que sur le plan technique. Le relâchement des contraintes géographiques qui existaient dans l'ancien plan de sondage a permis de pousser plus loin l'optimisation tandis que le fait de garder le critère de la contiguïté au premier niveau de stratification produira des unités qui se prêteront bien à la mise à jour de l'échantillon plus tard dans la décennie et facilitera la planification de la tâche des intervieweurs. En outre, dans l'ancien plan de sondage, il était probable qu'une strate AR serait entièrement confiée à un seul intervieweur, de sorte qu'il n'était pas possible de cerner la part de la variance totale qui était attribuable à la variance de réponse. Dans la mesure où, dans chaque strate, la tâche des intervieweurs est déterminée en fonction de critères géographiques et les strates secondaires sont formées selon des critères qui ne sont pas géographiques, on obtient dans le nouveau plan de sondage un recoupement des strates et des tâches des intervieweurs qui n'entraîne pas une augmentation du coût de la collecte des données et permet de mettre en évidence la part de la variance totale qui est attribuable à la variance de réponse.

A limiting factor in the stratification algorithm was that the number of stratification units should not exceed 300, after which convergence of the algorithm slowed substantially. To deal with this, the largest SRUs were subdivided into major parts based on natural criteria as suggested by the geography, prior to the optimal stratification. Cities subdivided in this fashion were Montreal (7 parts), Toronto (5), Winnipeg (2), Calgary (2), Edmonton (2), and Vancouver (3).

In non-CMA but Census-Tracted SRUs, a single level non-geographic stratification was carried out using the stratification algorithm. In the non-tracted SRUs, formation of one to three geographic strata was done manually without any attempt at optimality.

Table 1 presents study results for two SRUs – Ottawa and Quebec City – and compares efficiencies of the geographic strata used in the old design with those of optimal two-level strata. This is done using 1971 Census data. Percent reductions in the first stage variance due to stratification, calculated at the time of the 1981 Census, indicate the largest improvements are under the optimal stratification for income and rented dwellings. The marginal gains for other characteristics, including employed and unemployed, point to the strength and robustness of the simple, but deep, geographic stratification in the old design.

Un facteur limitatif de l'algorithme de stratification était que le nombre d'unités de stratification ne devait pas dépasser 300, car au delà de ce point la vitesse de convergence de l'algorithme diminuait de façon considérable. On a donc, avant la stratification optimale, subdivisé les plus grandes UAR en principales parties en fonction de divisions géographiques naturelles. Les villes ainsi divisées étaient Montréal (7 parties), Toronto (5), Winnipeg (2), Calgary (2), Edmonton (2) et Vancouver (3).

Dans les UAR qui étaient divisées en secteurs de recensement mais qui n'étaient pas des régions métropolitaines de recensement, on a procédé à un seul niveau de stratification selon des critères non géographiques au moyen de l'algorithme de stratification. Dans les UAR qui n'étaient pas divisées en secteurs de recensement, on a formé une à trois strates géographiques à la main, sans souci d'optimalité.

Le tableau 1 présente les résultats relatifs à deux UAR, Ottawa et la ville de Québec, d'une étude comparant, au moyen des données du recensement de 1971, l'efficacité de la stratification géographique utilisée dans l'ancien plan de sondage à celle de la stratification optimale à deux niveaux employée dans le nouveau plan. Les réductions en pourcentage de la variance au premier degré attribuable à la stratification, calculées au moment du recensement de 1981, montrent que les plus grandes améliorations apportées par la stratification optimale concernaient le revenu et les logements loués. Le fait que les gains n'aient été que marginaux pour ce qui est des autres caractéristiques, dont le nombre de personnes occupées et le nombre de personnes en chômage, donne une idée de la force et de la robustesse de la stratification géographique, simple mais détaillée, de l'ancien plan de sondage.

Table 1
Percentage Reduction in First Stage Variance Due to Stratification
Old vs. New Methods

Variable	Stratification Method	
	Méthode de stratification	
	Old	New
	Ancienne	Nouvelle
Total employed – Nombre de personnes occupées	9.1	12.6
Employment income – Revenu d'emploi	18.1	30.4
Secondary education – Études secondaires	39.4	42.1
Population 15+ – ans et plus	9.2	12.6
Population 15-24 – ans	12.9	17.6
Population 55+ – ans et plus	25.3	29.7
Total dwellings – Nombre total de logements	28.5	33.1
Dwellings rented – Nombre de logements loués	20.9	28.8
1 person households – Ménages d'une personne	33.7	38.4
2 person households – Ménages de deux personnes	27.5	29.6

Tableau 1
Réduction en pourcentage de la variance au premier degré attribuable à la stratification
Comparaison entre l'ancienne et la nouvelle méthode de stratification

Variable	Stratification Method	
	Méthode de stratification	
	Old	New
	Ancienne	Nouvelle
Agriculture ¹	5.9	3.9
Forestry/fishing ¹ – Exploitation forestière/pêche ¹	3.1	2.4
Mining ¹ – Mines ¹	4.8	3.0
Manufacturing – Secteur manufacturier	23.5	23.1
Construction – Bâtiment	11.9	11.2
Transport	4.2	6.4
Services	14.5	19.8
Unemployed ¹ – Chômeurs ¹	7.1	9.7

¹ Characteristics not used in optimization for new method. – Variables n'ayant pas servi à l'optimisation dans la nouvelle méthode

(b) Stratification of the Apartment Frame

To belong to the apartment frame, an apartment building must have at least 30 units and five floors of living quarters. The extent of stratification of the apartment frame varies from city to city, depending principally on the number of apartment buildings. Montreal and Toronto each have four geographic strata, some of which are further sub-stratified into size strata. In Toronto the geographic strata are further divided into three size strata, corresponding to apartment buildings with less than 100 dwellings, apartment buildings with between 100 and 200 dwellings, and those with over 200 units. In Montreal only two size strata were formed, with 100 dwellings being the boundary point between strata. The SRU of Hamilton has both geographic strata, (for the cities of Hamilton and Burlington respectively) and two size strata. Five other medium sized SRUs – Ottawa, Winnipeg, Calgary, Edmonton and Vancouver – have two size strata.

The remaining SRUs which had an apartment frame were Halifax, Quebec City, Hull, Oshawa, Windsor, London, Kitchener, St. Catharines, Saskatoon, and Victoria. For these cities stratification of the apartment frame was not considered necessary due to the smaller number of apartments.

In addition to the above stratification, a further implicit geographic stratification of apartment buildings on the frame at the time of the redesign was achieved by sorting the apartment frame units by the primary area frame stratum in which they were physically located. (Apartment building updates after the redesign are merely added to the bottom of the list for each stratum, and hence are not subject to this implicit stratification.) The numbering of apartment frame units incorporates the primary area stratum identification to facilitate placement of apartment frame selections with nearby area frame selections when forming assignments.

2.4 Stratification and Delineation of Primary Sampling Units in Non Self Representing Areas

2.4.1 General Considerations

The following considerations were used in the choice of methods for stratification and delineation of Primary Sampling Units in NSR areas.

- (i) **Number and size of PSU:** The Primary Sampling Unit should be relatively compact geographically, with a sample yield of 55-60 dwellings, equivalent to an interviewer's assignment. There should be two to three selected per stratum.

b) Stratification de la base d'appartements

Pour être inscrit dans la base d'appartements, un immeuble doit compter au moins 30 unités et avoir cinq étages de logements. Le nombre de strates formées dans la base d'appartements varie d'une ville à une autre, le principal critère étant le nombre d'immeubles d'appartements. Quatre strates ont été créées à Montréal et à Toronto respectivement selon un critère géographique, dont certaines ont ensuite été subdivisées selon un critère de taille. À Toronto, les strates géographiques ont été divisées en 3 strates formées selon la taille : une strate pour les immeubles de moins de 100 logements, une strate pour les immeubles de 100 à 200 logements et une strate pour les immeubles de plus de 200 logements. À Montréal, on n'a formé que 2 strates selon la taille, la ligne de démarcation étant 100 logements. L'UAR de Hamilton compte 2 strates géographiques (les villes de Hamilton et de Burlington) et en 2 strates formées selon la taille. On a fait 2 strates selon la taille aussi dans cinq autres UAR de taille moyenne, Ottawa, Winnipeg, Calgary, Edmonton et Vancouver.

Les autres UAR pour lesquelles on a constitué une base d'appartements étaient Halifax, la ville de Québec, Hull, Oshawa, Windsor, London, Kitchener, St. Catharines, Saskatoon et Victoria. Ces bases n'ont pas été stratifiées davantage vu le petit nombre d'appartements qu'elles contenaient.

En plus de la stratification décrite ci-dessus, on a procédé à une autre stratification géographique implicite des immeubles d'appartements qui faisaient partie de la base au moment du remaniement du plan de sondage en classant les unités de la base d'appartements selon la strate primaire de la base aréolaire dans laquelle ils étaient situés. (Les nouveaux immeubles d'appartements sont simplement rajoutés au bas de la liste dressée pour chaque strate, et par conséquent ne font pas l'objet de la stratification implicite). On inclut dans le numéro attribué aux unités de la base d'appartements les chiffres identifiant la strate primaire de la base aréolaire pour pouvoir grouper plus facilement dans la tâche d'un intervieweur les unités choisies dans la base d'appartements et les unités voisines choisies dans la base aréolaire.

2.4 Stratification et délimitation des UPE dans les secteurs non autoreprésentatifs

2.4.1 Critères généraux

On s'est basé sur les critères suivants pour choisir les méthodes de stratification et de délimitation à appliquer aux unités primaires d'échantillonnage dans les secteurs NAR.

- (i) **Nombre d'UPE et taille :** Une unité primaire d'échantillonnage (UPE) devrait être relativement compacte du point de vue géographique et avoir un rendement de l'échantillon de 55 à 60 logements, ce qui équivaut à la tâche d'un intervieweur. On devrait prélever 2 ou 3 UPE par strate.

(ii) **Stratification should be geographic within Economic Regions:** Robustness of stratification was a key consideration in the choice of stratification variables. It was also the principal motivation for deciding to have geographically contiguous and compact strata, since it was reasoned that areas near one another geographically had a better chance of retaining this homogeneity than did non-contiguous areas. Another motivation for geographic strata was to ensure a better geographic spread of the sample. This would generally tend to improve the capability for producing small area estimates for user defined domains.

(iii) **Within PSU heterogeneity:** In the old LFS design, strata were formed using Census Enumeration Areas (EA) as stratification units, and then PSUs (which were groups of EAs) were delineated within strata. This sequence permitted greater optimality to be introduced in delineating PSUs, since PSUs could be made as similar as possible to the *known* characteristics of the stratum, thereby reducing the between-PSU component of sampling variance. Conceptually, making PSUs as similar as possible to the stratum or to each other, is equivalent to increasing the within-PSU heterogeneity as much as possible. The delineation of heterogeneous PSUs within strata has been retained in the new design.

Bearing in mind the above considerations, and on the basis of studies carried out during the research phase of the redesign, the adopted design features explicit urban and rural stratification in Economic Regions where the population of area type is sufficient to permit separate strata. Roughly 50% of the Economic Regions (34 out of 71) qualify for this type of **urban-rural** design. In 31 of the remaining ER, a mixed design exists where primary strata within Economic Regions consist of both rural and urban areas. The mixed design is similar to the old design in the NSR areas.

Then there are some special cases. Economic Region 590 in Northern Ontario is unique, and is divided into two strata using the proportion of the labour force engaged in mining as the stratification variable. The remaining five economic regions do not contain any NSR areas. Four of them are in northerly parts of Quebec, Manitoba, Saskatchewan and British Columbia (which apart from some SRUs were designated remote), and the other is entirely Self Representing (Economic Region 67- the city of Winnipeg).

(ii) **Stratification géographique dans les RE :** Un des plus importants buts, que le choix des variables de stratification devait permettre d'atteindre, était la robustesse de la stratification. C'était principalement à cette fin aussi qu'il a été décidé que les strates seraient contiguës et compactes au plan géographique, en vertu du raisonnement selon lequel les régions voisines l'une de l'autre ont de meilleures chances de demeurer homogènes que les régions non contiguës. Le critère géographique a aussi été retenu parce qu'il assurerait une meilleure dispersion de l'échantillon, ce qui en général permet d'être davantage en mesure de produire des estimations régionales pour des domaines définis par les utilisateurs.

(iii) **Hétérogénéité à l'intérieur des UPE :** Dans l'ancien plan de sondage de l'EPA, on formait les strates en prenant d'abord les secteurs de dénombrement (SD) du recensement comme unités de stratification puis en délimitant les UPE (qui étaient des groupes de SD) à l'intérieur des strates. Cet ordre permettait de délimiter les UPE de façon plus optimale, car on pouvait faire en sorte que les caractéristiques de la population des UPE se rapprochent le plus possible des caractéristiques *connues* de la population de la strate et ainsi réduire la partie de la variance d'échantillonnage attribuable aux UPE. Théoriquement, si l'on s'arrange pour qu'il y ait le plus de similitude possible entre les UPE et la strate ou entre les UPE les unes par rapport aux autres, cela revient à augmenter l'hétérogénéité des UPE le plus possible. La délimitation d'UPE hétérogènes à l'intérieur des strates est un critère qui a été retenu dans le nouveau plan de sondage.

En tenant compte des critères énoncés ci-dessus et en se fondant sur les résultats d'études effectuées pendant la période de recherche qui a précédé le remaniement du plan de sondage, on a adopté un plan selon lequel les RE étaient explicitement stratifiées en régions rurales et en régions urbaines partout où la population était suffisante pour permettre de former des strates distinctes. Environ 50% des régions économiques (34 sur 71) ont pu se prêter à ce plan **urbain-rural**. Dans 31 des RE restantes, on a appliqué un plan mixte selon lequel les strates primaires dans les RE sont composées à la fois de régions rurales et de régions urbaines. Ce plan mixte ressemble à l'ancien plan pour les secteurs NAR.

Il y a ensuite quelques cas spéciaux. La RE 590 dans le nord de l'Ontario est unique en son genre, en ce sens qu'elle est divisée en 2 strates selon une variable de stratification qui est la proportion de la population active travaillant dans les mines. Les 5 RE restantes ne contiennent pas de secteurs NAR. Quatre d'entre elles sont situées dans le nord d'une province, en l'occurrence le Québec, le Manitoba, la Saskatchewan et la Colombie-Britannique, et, exception faite de quelques UAR, ont été classées dans la catégorie des régions éloignées, tandis que la dernière est entièrement autoreprésentative (RE 67, la ville de Winnipeg).

2.4.2 Urban-Rural Design

Within the urban portion of the Economic Region, individual urban centres are generally taken as PSUs although occasionally small urban centres are combined. In most Economic Regions there is only sufficient urban population to permit one urban stratum, whereas in a few cases the urban portion of the Economic Region is further stratified geographically.

Where the population of the rural portion of the Economic Region permits it (about 40% of cases), there is optimal sub-stratification. The 1981 Census enumeration area was used as the stratification unit in all provinces except Quebec and Ontario. The Economic Regions are larger in these provinces, with typically over 300 rural EAs per Economic Region, and due to the slow convergence of the stratification algorithm with this many units, rural Census Sub-divisions were taken as stratification units. The strata are geographically contiguous and compact strata, and are based on the 16 Census characteristics described earlier. (The number of urban and rural strata per Economic Region is given in Appendix E.)

The delineation of PSUs within rural strata is conceptually very similar to stratification. The only difference is that in stratification we attempt to minimize the sums of squares of the geographic and non-geographic variables within each stratum, while in PSU formation, we want to minimize the sums of squares of the geographic variables (to obtain compact PSUs in order to reduce costs) and to maximize those of the non-geographic variables. The latter criterion results in PSUs that are as heterogeneous as possible in terms of characteristics, so that they are all properly representative of the stratum during sampling.

There is, however, a conflict between the desired compactness of the PSUs and their heterogeneity, due to the tendency of adjacent units to possess similar characteristics. The low marginal computing costs permitted three delineations per stratum, with centroid weights of 10, 15 and 20 relative to the other variables. The results of each delineation were then plotted by computer on a graph whose axes are the centroids. The best of the three delineations was chosen based on the quality of the stratification as measured by stratification indices, and by visual inspection of the graphs for compactness. A compactness index was also taken into consideration. In most cases, a centroid weight of 10 or 15 was chosen.

2.4.2 Plan urbain-rural

Dans la partie urbaine d'une région économique, on prend généralement chaque centre urbain pour en faire une unité primaire d'échantillonnage, quoiqu'il arrive à l'occasion qu'on rattache les petites régions urbaines aux centres urbains voisins. Dans la plupart des RE, la population urbaine ne permet de former qu'une strate urbaine, sauf dans un petit nombre de cas où la partie urbaine d'une RE est stratifiée une fois de plus selon des critères géographiques.

Lorsque la population de la partie rurale d'une région économique le permet, (environ 40% des cas) la sous-stratification est optimale. On a pris les secteurs de dénombrement (SD) du recensement de 1981 comme unités de stratification dans toutes les provinces sauf au Québec et en Ontario. Dans ces provinces, les RE sont plus grandes, chacune comptant généralement plus de 300 SD ruraux; étant donné que la vitesse de convergence de l'algorithme de stratification est moins grande lorsqu'il y a tant d'unités, on a pris comme unités de stratifications les subdivisions de recensement rurales. Les strates sont contiguës et compactes et sont formées en fonction des 16 caractéristiques de recensement décrites plus haut. Le nombre de strates rurales et urbaines par région économique est donné en annexe.

Du point de vue théorique, la délimitation des UPE dans les strates rurales ressemble à de la stratification. La seule différence est que, lorsqu'on fait de la stratification, on essaie de minimiser la somme des carrés des variables géographiques et non géographiques dans chaque strate, tandis que lorsqu'on forme des UPE, on veut minimiser la somme des carrés des variables géographiques (pour obtenir des UPE compactes et ainsi réduire les coûts) et maximiser celle des variables non géographiques. On obtient ainsi des UPE qui sont aussi hétérogènes que possible du point de vue des caractéristiques de la population et donc bien représentatives de la strate lorsqu'elles sont échantillonnées.

Les deux buts poursuivis (la compacité et l'hétérogénéité des UPE) sont cependant en contradiction, car les unités adjacentes possèdent généralement des caractéristiques semblables. Comme les frais marginaux de calcul étaient faibles, on a pu faire 3 délimitations par strate en attribuant aux centroïdes des poids de 10, 15 et 20 par rapport aux autres variables. On a ensuite représenté par ordinateur chaque délimitation ainsi obtenue sur un graphique dont les axes étaient les centroïdes, puis on a choisi la meilleure des trois délimitations en fonction de la qualité de la stratification telle que mesurée par les indices de stratification et de la compacité des unités telle que révélée par l'examen visuel des graphiques. On a également utilisé un indice de compacité. Dans la plupart des cas, on a retenu un poids de 10 ou de 15 pour les centroïdes.

2.4.3 Mixed Design

The mixed design exists in the roughly 50 percent of Economic Regions where there is insufficient urban or rural population to form separate strata. For two-thirds of these ERs, the combined urban and rural population is insufficient to permit sub-stratification. For the ERs with sufficient population for more than one mixed stratum, rural EAs and urban centres were used as stratification units, and geographically contiguous strata were formed using the 16 Census variables.

For PSUs in the mixed strata, an additional constraint is that the proportion of urban population must be approximately the same in each PSU. This constraint, which existed in the old LFS design, was retained on the grounds that urban and rural distinctions are important, particularly for estimation of such characteristics as persons employed in agriculture.

The constraint on the urban-rural ratio results in an implicit urban/rural stratification in the mixed design. It was achieved as follows. Each urban centre was divided into a number of parts such that the total number of urban parts within a stratum was equal (or double in some cases) to the number of PSUs in that stratum. The number of parts per urban centre was determined in proportion to the population of the centre.

The optimal stratification program was applied, considering each part of an urban centre as a distinct stratification unit, and by considering urban population as an extra stratification variable. The weight assigned to this variable was adjusted to obtain the most evenly balanced rural/urban distribution possible within each PSU, without unduly disrupting compactness and overall optimization. This was done by trial and error. In practice, a weight of 10 or 15 on urban population, relative to the other variables, produced satisfactory results. Notwithstanding these attempts, in some strata one or two PSUs received no urban component, largely due to the lack of urban centres in the vicinity.

2.4.3 Plan mixte

On a appliqué un plan mixte dans les quelque 50% des RE où la population urbaine ou rurale n'était pas suffisante pour qu'on puisse former des strates distinctes. Dans les deux tiers de ces RE, la population urbaine et rurale combinée ne pouvait pas être sous-stratifiée pour la même raison. Dans les RE où la population était assez élevée pour qu'on puisse créer plus d'une strate mixte, on a pris les SD ruraux et les centres urbains comme unités de stratification et on a formé des strates géographiques contiguës en se servant des 16 variables de recensement.

Il y avait une contrainte de plus pour les UPE des strates mixtes, à savoir que la proportion de la population urbaine devait être à peu près la même dans chacune d'entre elles. Cette contrainte, qui existait déjà dans l'ancien plan de sondage, a été maintenue parce que la distinction région urbaine-région rurale est importante, particulièrement si l'on veut estimer certaines caractéristiques comme le nombre de personnes ayant un emploi dans le domaine de l'agriculture.

Le ratio à respecter pour ce qui est de la population urbaine et de la population rurale a entraîné une stratification implicite des régions urbaines et rurales, qui a été réalisée de la façon suivante. Chaque centre urbain a été divisé en un certain nombre de parties de manière que le nombre total de parties urbaines dans une strate soit égal au nombre d'UPE dans cette strate (ou égal à deux fois ce nombre, dans certains cas). Le nombre de parties par centre urbain était proportionnel à la population du centre.

On a exécuté le programme de stratification optimale en considérant chaque partie d'un centre urbain comme une unité de stratification distincte et la population urbaine, comme une variable de stratification supplémentaire. Le poids attribué à cette variable a été rajusté de manière à obtenir la distribution rurale/urbaine la plus équilibrée possible dans chaque UPE sans trop nuire à la compacité et à l'optimisation globale. Ce but a été atteint par tâtonnements. On a obtenu des résultats satisfaisants en attribuant un poids de 10 ou de 15 à la population urbaine par rapport aux autres variables. Malgré cela, dans certaines strates, une ou deux UPE n'ont pas eu de partie urbaine, principalement parce qu'il n'y avait pas de centre urbain à proximité.

Chapter 3

Self Representing Areas

3.1 Introduction

As noted earlier, the Self Representing Units correspond to those urban centres which are sufficiently large to yield an expected sample of 50 dwellings. Stratification of the Self Representing areas was discussed in the previous chapter. Formation and selection of units, together with the mechanism for updating the sample, are described in this chapter.

3.2 Formation of Self Representing Units

The SR universe was created in two steps: (i) the identification of cities, towns or Census Agglomerations (CAs) large enough to qualify as SRUs, and (ii) the delineation of boundaries. The criterion for designating a city, town or CA as SRU was that it should yield a sample of 50 or more dwellings. This cut-off was taken based on the judgement by survey operations staff of what should constitute a minimum assignment size, supported by analyses showing generally higher per unit data collection costs for smaller assignments.

For the Census Metropolitan Areas (CMAs), 1981 Census boundaries were automatically adopted as SRU boundaries, due to the requirement to publish CMA estimates. For remaining SRUs, decisions were made by examining maps case by case with a preference being given to respecting the 1981 Census boundaries; the exception was where they included large tracts of rural areas that would have increased travel distances and rendered the two-stage sample design expensive. Such departures occurred most frequently with CAs, since their boundaries are defined in terms of complete Census Subdivisions (CSDs). In such cases, SRU boundaries were defined either by dropping some CSDs altogether, or by splitting the CSD and including only those Enumeration Areas (EAs) that were reasonably built-up and close to the urban core.

3.3 Sample Design for Area Frame

3.3.1 Choice of Clusters

The main thrust in defining clusters, the first stage units, was to make use of 1981 Census units and dwelling counts wherever possible. This avoided the cost of an independent field count, which was used in the last redesign. The choice of unit depended on both theoretical and operational considerations.

Chapitre 3

Secteurs autoreprésentatifs

3.1 Introduction

Comme nous l'avons vu, les unités autoreprésentatives (UAR) correspondent aux centres urbains suffisamment grands pour avoir un rendement prévu de l'échantillon de 50 logements. La stratification des secteurs AR a déjà été décrite dans le chapitre précédent. Dans celui-ci, il sera question de la formation et de la sélection des unités ainsi que du mécanisme de mise à jour de l'échantillon.

3.2 Formation des unités autoreprésentatives

L'univers des unités autoreprésentatives a été créé en deux étapes : la première a consisté à déterminer quelles villes ou agglomérations de recensement étaient suffisamment grandes pour être des UAR; la seconde a consisté à délimiter les unités. Le critère pour qu'une ville ou agglomération de recensement soit considérée comme une UAR était qu'elle ait un rendement de l'échantillon d'au moins 50 logements. Ce seuil a été établi en fonction de ce qui, de l'avis du personnel des opérations de l'enquête, devrait être la tâche minimum d'un intervieweur, opinion étayée par les résultats d'analyses qui ont montré que le coût par unité de la collecte des données était généralement plus élevé lorsque la tâche était inférieure à ce seuil.

Pour les régions métropolitaines de recensement, les limites du recensement de 1981 sont automatiquement devenues celles des UAR du fait qu'on est tenu de publier des estimations s'y rapportant. Pour les autres UAR, la décision a été prise cas par cas après examen des cartes. On a donné la préférence aux limites du recensement de 1981 sauf lorsqu'elles englobaient de trop vastes étendues rurales, car les distances à parcourir auraient été trop grandes et le coût du plan d'échantillonnage à deux degrés aurait été trop élevé. C'est surtout pour les agglomérations de recensement qu'il a fallu faire ce genre d'exception, car leurs limites reprennent celles de subdivisions du recensement (SDR) complètes. Dans ces cas-là, pour délimiter les UAR, on a laissé tomber certaines SDR entières tandis que, de certaines autres, on a conservé seulement les SD qui étaient suffisamment bâtis et qui se trouvaient près du centre urbain.

3.3 Plan de sondage pour la base aréolaire

3.3.1 Choix des grappes

Le principe directeur pour la définition des grappes, c'est-à-dire les unités du premier degré, a été d'utiliser autant que possible les unités et les chiffres sur les logements du recensement de 1981 afin d'éviter d'avoir à supporter le coût du dénombrement indépendant sur le terrain qu'il avait fallu assumer avec l'ancien plan de sondage. Le choix des unités a été déterminé en fonction de critères d'ordre théorique et opérationnel.

A key operational consideration was clearly the level of geographic detail for which dwelling counts could be obtained from the 1981 Census. The lowest level for which automated retrieval of Census data could be done was the block face for areas falling under the coverage of the Area Master File (AMF). The AMF areas, accounting for 76% of the SR universe, corresponded to built-up portions of CMAs and the larger non-CMAs. The non-AMF portion of the SR universe was made up of the fringe portions of the CMAs and larger CAs, and smaller urban centres (comprising 13% and 11% of the SR universe, respectively). In non-AMF areas, the EA is the lowest unit for which Census counts can be retrieved by automation.

An alternative to using counts at the EA level for the non-AMF areas was to manually access the EA maps and visitation records used by Census enumerators to break down the counts into smaller units such as blocks and block faces. Given the availability of data for block faces, blocks and EAs, all three were examined as alternative choices for clusters. The block face was ruled out as being too small a unit which would have required frequent combining (e.g., over 20% had five or fewer dwellings).

From a sampling-theoretic point of view, for the case of an up-to-date design, there was not much of a choice between the EA and block: empirical results showed that variances for a two-stage design differed very little for the EA versus the block as a first stage unit (Choudhry, Lee and Drew, 1985). Hence, the choice was guided more by operational considerations of redesign and of robustness of the sample in the event of growth.

For the Area Master File areas, blocks were adopted as units since the availability of the counts in machine readable form from the Census database permitted a highly automated low cost redesign, the splitting or combining of blocks being required in less than 5% of cases. Moreover, the potential was recognized for linking blocks from one Census to the next due to their relative stability as geostatistical units. This permitted low cost quinquennial updating of the sample. In the fringe (non-AMF) portions of the larger urban areas, EAs were generally taken as the sampling unit. This was done on the grounds that in these areas with potential for large growth, an EA based design, due to the larger size of the units, would be less susceptible to deterioration of the sampling efficiency.

Finally for the non-AMF urban areas, the block was generally chosen as the sampling unit, and dwelling counts were obtained by manually accessing the visitation records. Although adoption of the EA would have required less initial work in designing the sample, over time this would have been offset by higher listing and list maintenance costs associated with EAs. Another important consideration for the choice of block over EA as the sampling unit in these SRUs was the design

Le niveau de détail géographique auquel il était possible d'obtenir des chiffres sur les logements dans la base de données du recensement de 1981 a certainement été un critère opérationnel important. Le plus faible niveau auquel on pouvait extraire automatiquement des données de recensement pour les régions représentées dans le fichier principal des régions (FPR) était le côté d'îlot. Les régions représentées dans le FPR (76% de l'univers AR) étaient les zones bâties des RMR et les grandes villes autres que les RMR. La partie de l'univers AR non représentée dans le FPR était composée des banlieues des RMR et des grandes agglomérations de recensement ainsi que des petits centres urbains (13% et 11% de l'univers AR respectivement). En ce qui concerne les régions non représentées dans le FPR, le SD est la plus petite unité pour laquelle il est possible d'extraire automatiquement des chiffres du recensement.

Il existait une autre possibilité que celle de prendre les chiffres au niveau des SD pour les régions non représentées dans le FPR, et c'était de se servir des cartes des SD et des registres des visites utilisés par les recenseurs pour ventiler les chiffres selon de plus petites unités, notamment les îlots et les côtés d'îlot. Comme on avait accès à des données sur les côtés d'îlot, les îlots et les SD, on a étudié les trois possibilités qu'ils représentaient pour la formation des grappes (unités de premier degré). Les côtés d'îlot ont été éliminés parce qu'ils auraient donné lieu à des unités trop petites qu'il aurait fallu grouper dans trop de cas (plus de 20% des côtés d'îlot ne comptaient pas plus de 5 logements).

Du point de vue théorique, pour un échantillon à deux degrés à jour, le choix était difficile à faire entre les SD et les îlots comme unités du premier degré, car, selon les résultats empiriques, les variances différaient peu selon qu'on prenait les uns ou les autres (Choudhry, Drew et Lee 1985). On a donc fondé la décision sur des critères d'ordre opérationnel, à savoir lequel des plans de sondage serait plus facile à remanier et lequel des deux échantillons serait plus robuste face à la croissance de la population.

Pour les régions représentées dans le FPR, on a pris les îlots comme unités, car le fait d'avoir accès aux chiffres sous forme ordinaire dans la base de données du recensement permettrait de remanier le plan de l'échantillon presque entièrement automatiquement et à peu de frais, moins de 5% des îlots ayant besoin d'être scindés ou groupés. En outre, on pensait qu'il serait possible de faire le lien entre les îlots d'un recensement à l'autre en raison de leur stabilité relative en tant qu'unités géostatistiques, ce qui permettrait de faire une mise à jour quinquennale de l'échantillon à peu de frais aussi. Pour les banlieues des grandes villes (non représentées dans le FPR), on a généralement pris les SD comme unités d'échantillonnage : la population de ces secteurs étant susceptible de s'accroître beaucoup, la grande taille des unités empêcherait probablement une baisse de l'efficacité du plan d'échantillonnage de se produire.

Enfin, pour les régions urbaines non représentées dans le FPR, l'unité d'échantillonnage choisie a généralement été l'îlot, et on a obtenu les chiffres sur les logements en consultant les registres des visites. La conception de la base de sondage aurait demandé moins de travail au départ si on avait pris les SD, mais alors ce gain aurait été plus que compensé par le coût élevé de l'établissement et de la mise à jour des listes. Il y a un autre critère important qui a joué un rôle dans le choix des îlots plutôt que des SD comme unités

constraint that six or a multiple of six clusters be selected per SRU (for equal representation of all rotation groups). Within the adoption of EAs as clusters, the smaller sizes of these SRUs would have yielded unacceptably high first stage sampling rates.

3.3.2 Formation of Clusters

- (a) **AMF Areas:** Formation of clusters in the AMF areas was carried out in five steps as described below:
 - (i) **Geographic File:** Geography Division created an area file for retrieval of Census data at the block face level. The area file contained the following for each block face: the block face, block and FED-EA numbers; the Universal Transverse Mercator (UTM) co-ordinates; street name and address range; and "pointers" linking the geographic entity to Census data on the Census database.
 - (ii) **Retrieval of Census Data:** Census Operations Division used the area file to retrieve the following Census data by block face:
 - counts of private occupied dwellings, broken down by high rise apartment dwellings versus all other dwelling types; and
 - counts of inmates and other residents of collective dwellings.
 - (iii) **Treatment of Exclusions:** All records belonging to an Indian reserve EA or to an EA belonging to the Special Area frame were dropped from the file.
 - (iv) **Unduplication of Area and Apartment Frames:** Adjustments were required to the Census counts of apartment and non-apartment dwellings to account for conceptual difference between the LFS and Census definitions of apartment (buildings with five floors and less than 30 units are classified as apartments by the Census but not by the LFS). This involved plotting LFS apartment clusters on Census blocks.
 - (v) **Formation and Mapping of Clusters:** The desired range for number of (non-apartment) dwellings in a cluster was set at 20-100. The preliminary computer-generated clusters corresponded to blocks, if the dwelling counts were in this range. The larger blocks were split if two clusters satisfying the size constraints could result from treating the block faces as separate clusters. Also, high and low (or zero) count blocks requiring manual splitting or combining were separately listed. The block face composition of clusters formed by the splitting and combining was subsequently data-captured to produce a machine readable file giving the block face composition of all clusters.

d'échantillonnage dans ces UAR, et c'est que le nombre de grappes formées dans chaque UAR devait être de six ou un multiple de six (afin que tous les groupes de renouvellement soient également représentés). Si on avait pris les SD comme grappes dans ces petites UAR, les fractions de sondage au premier degré d'échantillonnage auraient été trop élevées.

3.3.2 Formation des grappes

- a) **Régions représentées dans le FPR:** La formation des grappes dans les régions représentées dans le FPR a été faite en cinq étapes, décrites ci-après.
 - (i) **Fichier géographique :** La Division de la géographie a créé un fichier aréolaire pour l'extraction des données de recensement au niveau du côté d'îlot. Pour chaque côté d'îlot, il y avait dans le fichier aréolaire : le numéro du côté d'îlot, de l'îlot, de la CEF ainsi que du SD; les coordonnées de la projection transverse de Mercator (PTM); le nom de la rue et la fourchette d'adresses; des indicateurs permettant de faire le lien entre l'entité géographique et les données dans la base du recensement.
 - (ii) **Extraction des données de recensement:** La Division des opérations du recensement s'est servie du fichier aréolaire pour extraire les données de recensement suivantes par côté d'îlot :
 - nombre de logements privés occupés, répartis selon deux catégories : les logements dans les grands immeubles d'appartements et les autres logements;
 - nombre de détenus et autres résidents de logements collectifs.
 - (iii) **Traitement des exclusions :** Tous les enregistrements se rapportant à un SD qui correspondait à une réserve indienne ou qui faisait partie de la base des secteurs spéciaux ont été retirés du fichier.
 - (iv) **Non chevauchement de la base aréolaire et de la base d'appartements :** Il a fallu modifier les chiffres du recensement sur les appartements et les autres types de logements parce que la définition d'un appartement n'est pas la même pour le recensement et pour l'EPA (un immeuble à cinq étages comptant moins de 30 unités est considéré comme un immeuble d'appartements aux fins du recensement mais pas de l'EPA). Pour cela, on a représenté les grappes d'appartements graphiquement sur les îlots de recensement.
 - (v) **Formation et représentation des grappes sur les cartes :** On a fixé à 20 et 100 le nombre minimum et maximum de logements (autres que les appartements) qu'on voulait avoir dans une grappe. Les grappes provisoires générées par ordinateur correspondaient à des îlots, à condition que le nombre de logements tombe dans cette fourchette. On a subdivisé les grands îlots si, en traitant les côtés d'îlot comme des grappes différentes, on pouvait obtenir deux grappes répondant au critère de taille. En outre, on a mis sur une liste distincte les îlots dans lesquels le nombre de logements était trop élevé ou trop faible (voire nul), et qui donc devaient être subdivisés ou groupés manuellement. On a ensuite introduit dans l'ordinateur les données sur la composition par côté d'îlot des grappes obtenues par subdivision ou groupement d'îlots afin d'avoir un fichier informatique indiquant la composition de toutes les grappes par côté d'îlot.

Clusters were numbered serially within strata (also called subunits), and were manually plotted on 1981 Census Tract maps. Generally there were three to five Census Tracts per stratum.

- (b) **Non-AMF Areas:** For the 107 non-AMF Self Representing Units, cluster formation was carried out using 1981 Census Visitation Records (VRs) and accompanying Enumeration Area Maps. The Visitation Records are a basic working document of Census Representatives on which these persons list separately: (i) private occupied dwelling units, (ii) collective dwellings, and (iii) unoccupied dwelling units. Normally these lists contain civic addresses of the dwellings, but where no civic addresses exist, dwellings are often listed by the name of the occupant. Accompanying each visitation record is an Enumeration Area map, on which the Census Representatives plot the location of individual dwellings listed on the VR, and indicate by arrows the path followed in canvassing the EA.

The first step in cluster formation consisted of using these documents to indicate dwelling counts by block and block face on urban place maps of the SRUs. Other salient features were also plotted on the place maps, such as EA boundaries, and natural physical features such as railway tracks that could be used to sub-divide large counts. Collective dwellings and apartments with 30 or more units were also plotted, with descriptions in the margins. For rural EAs on the fringes of the SRUs, less detailed counting based on physical features was carried out. When counting was not possible – for example, where the EA map was of poor quality – the EA was left intact if its dwelling count was acceptable for a cluster. Otherwise, these EAs were sent for field counting, which was done for only 3% of the EAs.

After counts were transferred to the maps, clusters were delineated using criteria similar to those used in AMF areas; thus, clusters should be blocks, combinations of blocks, or block faces, with counts of 20-100 dwellings. Separately plotted apartment buildings were generally taken as clusters by themselves. As clusters were formed they were delineated on the maps, and the dwelling counts were entered on control sheets for data capture for computerized selection of clusters.

On a numéroté en série les grappes dans les strates (appelées aussi sous-unités) et on les a représentées graphiquement sur les cartes des secteurs de recensement de 1981. Il y avait généralement de 3 à 5 secteurs de recensement par strate.

- (b) **Régions non représentées dans le FPR :** Pour les 107 UAR des régions non représentées dans le FPR, on a formé les grappes à l'aide des registres des visites du recensement de 1981 et des cartes des secteurs de dénombrement qui accompagnaient ces derniers. Les registres des visites sont des documents de travail de base des recenseurs sur lesquels ceux-ci inscrivent séparément la liste (i) des unités de logements privés occupés; (ii) des logements collectifs; (iii) des unités de logements inoccupés. Normalement, les logements sont identifiés par leur adresse de voirie; s'il n'y a pas d'adresse de voirie, ils le sont souvent par le nom de l'occupant. Chaque registre des visites est accompagné d'une carte du secteur de dénombrement, sur laquelle le recenseur indique l'emplacement de chaque logement inscrit dans le registre des visites et, au moyen de flèches, l'itinéraire qu'il suit lorsqu'il parcourt le secteur.

La première étape de la formation des grappes a consisté à se servir de ces documents pour indiquer le nombre de logements par îlot et par côté d'îlot sur les cartes des régions urbaines des UAR. On a indiqué d'autres éléments importants sur ces cartes, comme les limites des SD, ainsi que les particularités physiques telles les voies ferrées dont on pourrait se servir pour subdiviser les secteurs à population trop élevée. On a indiqué de façon distincte les logements collectifs et les immeubles d'appartements comptant 30 unités ou plus, et on a ajouté dans les marges des descriptions s'y rapportant. Pour ce qui est des SD ruraux à la limite des UAR, on a indiqué des données moins détaillées en fonction des divisions physiques plutôt que des unités de logement. Là où il n'était pas possible d'indiquer de chiffres, notamment parce que la carte du SD était très mauvaise, on a laissé le SD tel quel à condition que le nombre de logements permette de former une grappe. Autrement, on a demandé un dénombrement sur le terrain du SD, ce qui n'a été fait que pour 3% des SD.

Après avoir transféré les chiffres sur les cartes, on a délimité les grappes selon des critères semblables à ceux utilisés pour les régions représentées dans le FPR : les grappes devraient être des îlots, des groupes d'îlots ou des côtés d'îlots et comporter entre 20 et 100 logements. Les immeubles d'appartements indiqués de façon distincte ont généralement constitué des grappes à eux seuls. Au fur et à mesure que les grappes ont été formées, on les a délimitées sur les cartes et on a inscrit les nombres de logements sur les feuilles de contrôle pour la saisie des données en vue de la sélection automatique des grappes.

3.3.3 Choice of the Selection Method

From each stratum, six (or in special cases 12) clusters were selected using the method due to Rao, Hartley and Cochran (1962). Considerations which led to the retention of this method are the same as those that led to its initial adoption at the time of the 1973 redesign, and are as follows:

- (a) **Simplicity in Updating:** Since changes in the size measures are inherent in any continuous survey based on area sampling methods, the adaptability of the selection method to a simple procedure of revision of the selection probabilities is an important criterion for the choice of a selection method in the initial design. For many complex selection methods involving two or more selections, the revision of initial selection probabilities requires complex treatments due to their diverse joint inclusion probabilities. The Rao-Hartley-Cochran (RHC) method, like the method of selecting one unit per stratum, can be easily adapted to simple revision of the selection probabilities.
- (b) **Number of Sample Clusters:** In rotating panel designs, it is desirable that each panel be a representative sub-sample of the population. The importance of this for the LFS is discussed further in Chapter 6.

For each rotation group to be fully representative, the existence of six rotation groups implies that the sample selection method should permit the selection of six clusters or a multiple of six clusters per stratum. Most of the unequal probability schemes suggested in the literature consider the selection of two units only, and in this sense would not be suitable for application in the SRUs of the LFS.

Another option, which would still permit sample updating, would have been to have smaller strata with only one selection per stratum, and with representativity of rotation groups being achieved only over collections of strata. However, this was ruled out as being less desirable due to problems in representativeness of rotation groups, in estimation of variance, and in using rotation groups to select representative sub-samples for other surveys.

- (c) **Flexibility for Sample Size Changes:** Since the LFS has undergone major sample redesigns only at ten year intervals, a critical design feature is the flexibility to incorporate changes in the sample size. Such changes may be dictated by external factors such as shifts in government priorities or changes in the information needs to be satisfied by the survey. Or they may simply be necessitated due to technological changes in data collection resulting in more cost efficient methods.

3.3.3 Choix de la méthode de sélection

On a choisi six (ou, dans des cas spéciaux, 12) grappes dans chaque strate par la méthode de Rao, Hartley et Cochran (1962). On a conservé cette méthode pour les mêmes raisons que celles qui avaient motivé son adoption lors du remaniement de 1973, c'est-à-dire :

- a) **Simplicité de la mise à jour :** Etant donné que dans toute enquête permanente fondée sur des méthodes de sondage aréolaire la taille des unités de sélection varie, le choix de la méthode de sélection à utiliser dans le plan de sondage initial dépend beaucoup de la possibilité de réviser par la suite, par un procédé simple, les probabilités de sélection. Pour beaucoup de méthodes de sélection complexes impliquant deux sélections ou plus, la révision des probabilités de sélection initiales nécessite des traitements complexes en raison des probabilités composées d'inclusion. La méthode de Rao, Hartley et Cochran (RHC), comme la méthode qui consiste à choisir une unité par strate, se prête bien à une révision simple des probabilités de sélection.
- b) **Nombre de grappes d'échantillon :** Dans les plans de sondage par panel avec renouvellement de l'échantillon, il est souhaitable que chaque panel soit un sous-échantillon représentatif de la population. Nous parlerons davantage de l'importance de ce critère pour l'EPA dans le chapitre sur le renouvellement de l'échantillon.

Comme il y a six groupes de renouvellement, pour que chacun soit parfaitement représentatif, il faut que la méthode de sélection de l'échantillon permette de choisir six grappes (ou un multiple de six) par strate. La plupart des méthodes de sélection avec probabilités inégales proposées dans les ouvrages spécialisés n'envisagent que la sélection de deux unités; on a donc estimé qu'on ne pouvait pas les appliquer aux UAR de l'EPA.

Une autre méthode, qui aurait aussi permis la mise à jour de l'échantillon, aurait consisté à former de petites strates et à ne faire qu'une sélection par strate, la représentativité des groupes de renouvellement étant obtenue sur plusieurs strates. Cette méthode a cependant été écartée parce qu'elle a été jugée moins intéressante du fait qu'elle posait des problèmes sur le plan de la représentativité des groupes de renouvellement, de l'estimation de la variance et de l'utilisation des groupes de renouvellement pour la sélection de sous-échantillons représentatifs pour d'autres enquêtes.

- c) **Modification de la taille de l'échantillon :** Comme le plan de sondage de l'EPA n'est remanié en profondeur que tous les dix ans, il est très important qu'il permette de modifier facilement la taille de l'échantillon. Ces modifications peuvent être rendues nécessaires par des facteurs externes, comme un changement des priorités de l'Etat ou l'évolution des besoins en information que l'enquête doit satisfaire, ou simplement par les progrès technologiques dans le domaine de la collecte des données qui entraînent la mise au point de méthodes plus économiques.

The RHC method can readily accommodate changes in the overall sample size, or the allocation of sample to stages, or both. This method has been adapted for the LFS to make it totally flexible both to changes in the overall sampling rate, and to changes in the number of random groups formed. This is discussed further in relation to sample updating.

3.3.4 Sampling Methodology

- (a) **Selection of Clusters:** The selection of clusters using the RHC method is as follows. For a given stratum, let N = the number of clusters, R = the stratum inverse sampling ratio (isr), and n = the number of random groups (where $n = 6$, or 12 in special cases). Then the N clusters are randomly partitioned into n groups, containing either $[N/n]$ or $[N/n] + 1$ clusters, where $[x]$ refers to the greatest integer less than or equal to x .

The expected number of dwellings for each group, j , is computed as

$$d_j = \frac{D_j}{R},$$

where D_j is the number of dwellings in group j ($j = 1, \dots, n$)

The cluster inverse sampling ratio R_{jk} was obtained for all clusters k in group j as

$$R_{jk} = \left[\frac{D_{jk}}{d_j} \right]$$

or

$$\left[\frac{D_{jk}}{d_j} \right] + 1,$$

where D_{jk} is the number of dwellings in cluster k of group j , so that

$$\sum_k R_{jk} = R.$$

Within each group, one cluster was selected with probability proportional to its size.

Since the total sample size for the stratum, say m , is fixed ($m = D/R$, where D is the number of dwellings in the stratum), clearly, increasing the number of clusters selected will imply fewer dwellings per selected cluster. Studies were carried out at the time of redesign to determine an optimal allocation of the sample between the two stages of sampling. A cost model assuming a personal visit for dwellings in their first month in the sample and telephone interviews in later months was developed. Based on a variance component analysis and the cost model, optimal within-cluster sample sizes (or density factors) were found to be five to six dwellings in strata with blocks as clusters, and nine to ten dwellings in strata with EAs as clusters (Choudhry, Lee and Drew, 1985).

La méthode RHC se prête facilement à la modification de la taille globale de l'échantillon ou à la répartition de l'échantillon selon le degré d'échantillonnage, ou aux deux. On a adapté la méthode à l'EPA de manière à lui donner une souplesse absolue qu'il s'agisse de modifier la fraction de sondage globale ou le nombre de groupes aléatoires formés. Ces questions sont traitées plus loin dans le contexte de la mise à jour de l'échantillon.

3.3.4 Méthodes d'échantillonnage

- (a) **Sélection des grappes :** La sélection des grappes par la méthode RHC se fait de la façon suivante. Dans une strate données, soit N = le nombre de grappes, R = la fraction de sondage inverse de la strate (fsi) et n = le nombre de groupes aléatoires (où $n = 6$ ou 12 dans les cas spéciaux). Les N grappes sont partitionnées au hasard en n groupes contenant $[N/n]$ ou $[N/n] + 1$ grappes, où $[x]$ est le plus grand nombre entier inférieur ou égal à x .

Le nombre prévu de logements dans chaque groupe, j , est calculé de la manière suivante

$$d_j = \frac{D_j}{R},$$

où D_j est le nombre de logements dans le groupe j ($j = 1, \dots, n$).

On obtient la fraction de sondage inverse des grappes R_{jk} pour toutes les grappes k d'un groupe j par la formule

$$R_{jk} = \left[\frac{D_{jk}}{d_j} \right]$$

ou

$$\left[\frac{D_{jk}}{d_j} \right] + 1,$$

de manière à ce que

$$\sum_k R_{jk} = R.$$

On choisit une grappe dans chaque groupe selon une probabilité proportionnelle à sa taille.

Comme la taille de l'échantillon pour la strate, soit m , est fixe ($m = D/R$, où D est le nombre de logements dans la strate), il est clair qu'une augmentation du nombre de grappes échantillonnées entraîne une diminution du nombre de logements par grappe choisie. Des études ont été réalisées au moment du remaniement du plan de sondage pour déterminer la répartition optimale de l'échantillon entre les deux degrés d'échantillonnage. En se fondant sur une analyse des composantes de la variance et sur un modèle de coûts supposant une visite sur place le premier mois qu'un logement participe à l'enquête et des interviews téléphoniques les mois suivants, on a trouvé que la taille optimale d'une grappe (ou facteur de densité) était de 5 ou 6 logements dans les strates où les îlots sont des grappes et de 9 ou 10 logements dans les strates où les SD sont des grappes (Choudhry, Lee et Drew 1985).

To ensure density factors in the desired ranges, with a choice of $n = 6$, a constraint was that the sample yields should be in the range of 25-35 dwellings for strata with a block-based design, and 40-50 dwellings for strata with an EA-based design. Occasionally, when strata exceeded the usual size constraints, the choice of $n = 12$ resulted in density factors closest to the desired values.

(b) Sampling within Clusters

Within selected clusters two random numbers, r_1 and r_2 , in the interval $[1, R_{jk}]$ are determined. The first of these r_1 , determines the number of occasions (i.e., six month periods) the selected cluster will remain in the sample before being replaced. The second, r_2 , determines the random start to be used in selecting the initial systematic sample of dwellings from the cluster. On each of the $(r_1 - 1)$ successive sampling occasions within the cluster, the random start is incremented by 1, with start 1 following start R_{jk} .

3.4 Sample Update Methodology

Under the old LFS design, during the period from 1975 to 1978, SRU design effects (i.e., the ratio of observed variances to expected variances under simple random sampling) had been increasing by 7-8% a year. In 1978, a program of sample updating was initiated, with sufficient intensity to arrest this deterioration during the period 1978-1982. The SRU update program was terminated coincident with the startup of work on the post-1981 Censal redesign. The SRU updating under the old design demonstrated the cost effectiveness of such updating as a means of maintaining desired levels of reliability for survey results.

The key features of the RHC method for sample updating are that after formation of random groups, sampling is carried out independently within each random group, and that only one unit is selected with PPS per group. This permits a Keyfitz (1951) update, where new probabilities of selection can be incorporated while maximizing the probability of retaining already selected units. Retention of already selected units is desirable, since it avoids the expense of initial creation and capture of dwelling lists associated with the selection of new clusters. During the sample updating from 1978-1982, originally selected clusters were retained 70% of the time.

The update methodology consists of the following steps:

- (i) Obtain new counts for existing clusters in a stratum, with cluster breakdowns in the case of excessive growth.

Pour s'assurer que les facteurs de densité tombent dans les limites voulues lorsque $n = 6$, on a posé une condition relativement à la taille au moment de la stratification, à savoir que le rendement de l'échantillon soit de 25 à 35 logements dans les strates où l'unité de base est l'ilot et de 40 à 50 logements dans les strates où l'unité de base est le SD. A l'occasion, lorsque la taille d'une strate dépassait la limite établie, on a obtenu des facteurs de densité plus près des valeurs voulues en prenant $n = 12$.

b) Échantillonnage à l'intérieur des grappes

Dans les grappes choisies, on détermine deux nombres aléatoires r_1 et r_2 , dans l'intervalle $[1, R_{jk}]$. Le premier de ces nombres, r_1 , représente le nombre de cycles (c.-à-d. de périodes de six mois) que la grappe choisie restera dans l'échantillon avant d'être remplacée et r_2 définit l'origine choisie au hasard qui servira au prélèvement du premier échantillon systématique de logements dans la grappe. A toutes les $(r_1 - 1)$ fois qu'on procède à l'échantillonnage dans la grappe, on ajoute 1 à l'origine choisie au hasard, l'origine R_{jk} étant toujours suivie de l'origine 1.

3.4 Méthode de mise à jour de l'échantillon

Lorsqu'on utilisait l'ancien plan de sondage, entre 1975 et 1978, l'effet du plan pour les UAR (c.-à-d. le ratio de la variance observée à la variance prévue selon l'échantillonnage aléatoire simple) augmentait de 7 à 8% par an. En 1978, un programme de mise à jour de l'échantillon a été institué et il a été suffisamment intensif pour qu'en 1982 cette détérioration ait cessé. Le programme de mise à jour de l'échantillon des UAR a pris fin au moment-même où l'on a amorcé le remaniement de l'EPA, après le recensement de 1981. La mise à jour des UAR dans l'ancien plan de sondage a montré que cette façon de maintenir le degré de fiabilité des résultats de l'enquête au niveau voulu était économique.

La principale caractéristique de la méthode RHC du point de vue de la mise à jour de l'échantillon est que, après la formation des groupes aléatoires, chacun de ces groupes est échantillonné indépendamment et seulement une unité y est choisie avec probabilité proportionnelle à la taille. Cette façon de procéder permet de faire une mise à jour par la méthode de Keyfitz (1951), selon laquelle il est possible d'introduire de nouvelles probabilités de sélection tout en maximisant la probabilité de garder les unités déjà choisies. La raison pour laquelle on veut conserver ces unités est que cela évite les frais de création et de saisie des listes de logements entraînés par la sélection de nouvelles grappes. Lors de la mise à jour effectuée entre 1978 et 1982, on a gardé 70% des grappes qui avaient été prélevées initialement.

La méthode de mise à jour comporte les étapes suivantes :

- (i) Obtenir les nouveaux chiffres pour les grappes faisant déjà partie de la strate, et les ventiler si la croissance a été excessive.

- (ii) Re-cluster the cases with excessive growth. Assign the new clusters randomly to groups such that the number of clusters per random group in the stratum remains constant.
- (iii) At this point random groups consist of say, N_1 old clusters retained at the time of updating, N_2 old clusters split into two or more new clusters at the time of updating, and N_3 new clusters formed by splitting of old clusters in the stratum.

The cluster $isr (R_{jk})$ is already defined for each of the $N_1 + N_2$ units existing in the old design. For the N_3 new units, define $R_{jk} = 0$.

Similarly, define the new cluster $isr R_{jk} = 0$ for the N_2 old clusters which were split up at the time of updating. For the $N_1 + N_3$ new clusters, define revised isr 's in the same manner in which the old isr 's were defined, but based on the revised dwelling counts.

Old and new probabilities of selection of clusters are given respectively by

$$P_{jk} = \frac{R_{jk}}{R},$$

and

$$P'_{jk} = \frac{R'_{jk}}{R}.$$

- (iv) A Keyfitz sample update is performed within each group using old and new probabilities of selection. The Keyfitz procedure incorporates the new probabilities of selection while maximizing the retention of already selected units.

If a cluster is retained, with $R_{jk} = R_{jk}$, then the within cluster sampling remains unaffected, both in terms of the number of occasions in which samples of dwellings will be drawn from the cluster, and in terms of the actual systematic samples used.

- (v) If a cluster is retained but $R_{jk} \neq R_{jk}$, then the length of time the cluster will remain in the sample is determined such that sampling can be restricted to previously unselected dwellings. Clustering of the systematic samples of dwellings drawn under the old and new cluster isr 's is then prevented by use of an algorithm which re-arranges the ordering of dwellings on the cluster list.

3.5 Sample Design for the Apartment Frame

A separate frame of apartment buildings exists for 18 of the largest SRUs, namely Montreal, Toronto, Ottawa, Hamilton, London, Winnipeg, Calgary, Edmonton, Vancouver, Halifax, Quebec City, Oshawa, Windsor, Kitchener, Saskatoon, Victoria, Hull and St. Catharines. Changes from the previous design include the addition of Hull and St. Catharines, due to an increased number of high rise apartment dwelling units reported in the 1981 Census.

- (ii) Former de nouvelles grappes là où la croissance a été excessive. Répartir les nouvelles grappes au hasard parmi les groupes de manière que le nombre de grappes par groupe aléatoire dans la strate demeure constant.
- (iii) A ce stade-ci, les groupes aléatoires sont composés de, disons, N_1 anciennes grappes qu'on a gardées lorsqu'on a fait la mise à jour, N_2 anciennes grappes dont chacune a été divisée en deux nouvelles grappes ou plus au moment de la mise à jour et N_3 nouvelles grappes formées lorsque les anciennes grappes de la strates ont été divisées.

La fsi de la grappe (R_{jk}) est déjà définie pour chacune des $N_1 + N_2$ unités de l'ancien plan de sondage. Pour les N_3 nouvelles unités, nous posons $R_{jk} = 0$.

De même, nous posons la nouvelle fsi de la grappe $R_{jk} = 0$ pour les N_2 anciennes grappes qui ont été divisées lors de la mise à jour et, pour les $N_1 + N_3$ grappes, nous définissons les nouvelles fsi de la même manière qu'avant la mise à jour mais en fonction des nouveaux chiffres sur les logements.

Les anciennes et les nouvelles probabilités d'échantillonnage des grappes sont définies respectivement par

$$P_{jk} = \frac{R_{jk}}{R},$$

et

$$P'_{jk} = \frac{R'_{jk}}{R}.$$

- (iv) On fait la mise à jour dans chaque groupe par la méthode de Keyfitz en se servant des anciennes et des nouvelles probabilités de sélection. La méthode de Keyfitz permet d'inclure les nouvelles probabilités d'échantillonnage tout en maximisant la possibilité de garder les unités déjà choisies.

Si une grappe est conservée, et que $R_{jk} = R_{jk}$, l'échantillonnage dans la grappe ne change pas, ni du point de vue du nombre de fois que des échantillons de logements seront prélevés de la grappe ni du point de vue des échantillons systématiques utilisés.

- (v) Si une grappe est conservée mais que $R_{jk} \neq R_{jk}$, la période pendant laquelle la grappe restera dans l'échantillon est déterminée de manière que l'échantillonnage soit limité aux logements qui n'ont jamais été choisis auparavant. Pour éviter que les échantillons systématiques de logements prélevés au moyen des anciennes et des nouvelles fsi de grappe soient groupés, on utilise un algorithme qui refait l'ordre des logements dans la liste des grappes.

3.5 Plan de sondage pour la base d'appartements

Il existe une base de sondage distincte pour les immeubles d'appartements dans 19 des plus grandes UAR, soit Montréal, Toronto, Ottawa, Hamilton, London, Winnipeg, Calgary, Edmonton, Vancouver, Halifax, Québec, Oshawa, Windsor, Kitchener, Saskatoon, Victoria, Hull et St. Catharines. La constitution de ce genre de base pour Hull et St. Catharines est un des changements apportés par rapport à l'ancien plan, les résultats du recensement de 1981 ayant révélé que le nombre d'unités de logements dans les grands immeubles d'appartements avait augmenté dans ces villes.

The apartment frame was initially adopted as a feature of the SRU design during the 1960s. Its principal benefit is to eliminate one of the major causes of growth in area frame clusters. Additionally it improves representation in the sample of both apartment and non-apartment dwellers, who differ with respect to certain characteristics of interest. Stratification of the apartment frame is described in Chapter 2.

The apartment frame is open-ended, and new apartment buildings that satisfy the criteria are added to the existing frame regularly, using data from Canada Post and the Central Mortgage and Housing Corporation. The apartment frame was continuously updated under the old design. As such, its redesign was principally a matter of adjusting coverage to correspond to changes in the boundaries of the SRUs: apartment buildings were deleted in areas dropped from SRU coverage.

For the added cities of Hull and St. Catharines, and for annexed portions of other cities, qualifying apartment buildings were initially identified by means of field checks. Another redesign operation included the plotting of apartment frame units on 1981 Census Tract maps. The identification of the block in which the apartment building was located was captured for use in the AMF areas in unduplicating the apartment and area frames.

Within each size-geography stratum, the apartment buildings (clusters) were selected systematically, with probability proportional to size achieved by applying a pre-specified density factor of three. Dwellings, selected at the second stage of sampling, were chosen systematically using the procedure described for the area frame clusters.

C'est pendant les années 60 qu'on a ajouté les bases d'appartements au plan de sondage pour les UAR. Leur principal avantage est que, grâce à elles, une des principales causes de la croissance des grappes dans la base aréolaire est éliminée. De plus, les personnes qui habitent dans un appartement et celles qui habitent dans un autre genre de logement, et qui diffèrent du point de vue de certaines caractéristiques à l'étude, sont ainsi mieux représentées dans l'échantillon. La stratification de la base d'appartements est décrite au chapitre 2.

La base d'appartements est ouverte, ce qui veut dire que les nouveaux immeubles qui répondent aux critères y sont ajoutés régulièrement, en fonction de renseignements provenant de la Société canadienne des postes et de la Société canadienne d'hypothèques et de logement. Comme la base d'appartements était continuellement mise à jour lorsque l'ancien plan de sondage était en vigueur, son remaniement a consisté principalement à modifier sa composition pour tenir compte des changements apportés aux limites des UAR; on a notamment retiré de la base des immeubles situés dans des régions qui ne faisaient plus partie de l'univers des UAR.

Pour les villes de Hull et de St. Catharines et les secteurs qui ont été annexés à d'autres villes, on a tout d'abord repéré les immeubles d'appartements pouvant figurer dans la base de sondage en faisant une vérification sur le terrain. On a ensuite indiqué les unités de la base d'appartements sur les cartes des secteurs de recensement de 1981. Enfin, on a saisi l'identification des îlots dans lesquels les immeubles d'appartements étaient situés pour faire en sorte que, pour les régions non représentées dans le FPR, les unités ne se retrouvent pas à la fois dans la base d'appartements et dans la base aréolaire.

Dans chaque strate définie selon un critère de taille et un critère géographique, on a choisi les immeubles d'appartements (grappes) par la méthode d'échantillonnage systématique avec probabilité proportionnelle à la taille en utilisant un nouveau facteur de densité dont la valeur était de 3. Au deuxième degré d'échantillonnage, on a choisi les logements par la même méthode d'échantillonnage systématique que celle qui a été appliquée aux grappes de la base aréolaire.

Chapter 4

Non-Self Representing Areas

4.1 Introduction

The Non-Self Representing areas are those lying outside the Self Representing areas, and comprise rural areas and small urban centres. As noted earlier, a multi-stage sample design is used in NSR areas, where the sample within each Economic Region or stratum has been concentrated into relatively few compact areas (Primary Sampling Units), each with a sample yield capable of supporting one interviewer. Such a clustered design was dictated by the low population density and the large land areas involved. Another influencing factor was that the area sampling methodology on which the survey is based requires interviewers to make personal visits to dwellings, at least during their first month in the sample. Some other considerations related to stratification and delineation of primary sampling units in NSR areas were given in Chapter 2.

The redesign led to four major changes in the survey design in NSR areas, namely (i) the extension of telephone interviewing to NSR areas, (ii) the elimination of a stage of sampling in rural areas, (iii) the adoption of a design featuring explicit rural/urban stratification wherever feasible, and (iv) a special design for Prince Edward Island. The rationale behind each of these changes is briefly discussed in section 4.2.

The method of sampling PSUs described in section 4.3 is the same for each type of strata, namely Rural, Urban and Mixed. The subsampling methods for each of these strata types are then discussed in sections 4.4-4.6. Finally, the treatment of the special area sample is described in the last section.

4.2 Rationale for Changes in the Design

4.2.1 Extension of telephone interviewing

Telephone interviewing for months two to six in the sample was introduced during the early 1970s in SR areas, primarily to reduce cost. However in NSR areas, all interviewing continued to be done in person. One of the concerns about telephone interviews was possible respondent resistance due to the nature of the data being collected and the prevalence of party lines. But in recognition of the immediate cost savings as well as the potential longer term gains from adoption of new telephone survey technologies, it was decided to test the feasibility of extending telephone interviewing to NSR areas.

Chapitre 4

Unités non autoreprésentatives

4.1 Introduction

Les secteurs non autoreprésentatifs sont les secteurs situés à l'extérieur des secteurs autoreprésentatifs et sont formés de régions rurales et de petits centres urbains. Comme nous l'avons vu auparavant, on utilise un plan de sondage à plusieurs degrés dans les secteurs NAR, l'échantillon prélevé dans chaque région économique ou strate y étant concentré en un nombre relativement faible de régions compactes (unités primaires d'échantillonnage ou UPE) qui ont chacune un rendement de l'échantillon correspondant à la tâche d'un intervieweur. Un tel plan avec grappes a été adopté en raison, premièrement, de la faible densité de la population et des vastes étendues de terres dans ces secteurs et, deuxièmement, de la méthode de sondage areolaire sur laquelle repose l'enquête et qui veut que les intervieweurs se rendent personnellement dans les logements au moins le premier mois de l'enquête.

Le chapitre 2, sur la stratification, traite de quelques autres critères qui sont entrés en ligne de compte relativement à la stratification et à la délimitation des unités primaires d'échantillonnage dans les secteurs NAR. Le remaniement du plan de sondage a entraîné quatre changements importants en ce qui a trait aux secteurs NAR : (i) les interviews téléphoniques ont été étendues à ces secteurs; (ii) un degré d'échantillonnage a été supprimé dans les régions rurales; (iii) une stratification rurale/urbaine explicite a été adoptée partout où cela a été possible et un plan spécial a été conçu pour l'Île-du-Prince-Édouard. La raison de ces changements est brièvement expliquée à la section 2.

La méthode d'échantillonnage des UPE décrite à la section 3 a été appliquée aux trois sortes de strates, à savoir les strates rurales, urbaines et mixtes. Les méthodes de sous-échantillonnage utilisées pour chacune de ces sortes de strates sont décrites dans les sections 4 à 6. Enfin, le traitement de l'échantillon dans les secteurs spéciaux fait l'objet de la dernière section.

4.2 Raison des changements apportés au plan de sondage

4.2.1 Interviews téléphoniques

Au début des années 70, on a commencé à pratiquer l'interview téléphonique auprès des ménages échantillonnés dans les secteurs autoreprésentatifs pendant les mois 2 à 6 principalement dans le but de réduire les coûts. Dans les secteurs NAR, on a continué de faire toutes les interviews sur place, car on craignait que les répondants ne s'opposent à l'interview téléphonique en raison de la nature des données recueillies et du nombre élevé de lignes partagées par plusieurs abonnés. Toutefois, compte tenu des économies immédiates que le recours à l'interview téléphonique ont permis de réaliser ainsi que des autres avantages susceptibles d'en découler à long terme, on a décidé d'étudier la faisabilité d'étendre cette technique aux secteurs NAR.

Field testing was carried out during 1982-84 on small portions of the actual LFS sample. Since the primary objective of the testing was to assess the data quality implications of the procedure, interviewer assignments were split between telephone and personal interviewing procedures. The test findings were that nonresponse rates were marginally lower for the telephone sample without any detectable difference in estimates for labour force characteristics.

The decision to extend telephone interviewing led to an increase of about 10 to 15% in assignment sizes to an average of 55 to 65 dwellings. Also, rotation numbers were assigned in rural areas at a higher stage, so that within secondary units all dwellings received the same rotation number. This cut down on the number of visits to the secondary units in months two to six in the sample.

4.2.2 Elimination of stage of sampling in rural areas

In the old design, as mentioned in Chapter 1, the rural sample within PSUs was selected in three stages: secondaries (called groups), clusters, and dwellings. The clusters corresponded to identifiable land areas containing up to 20 dwellings, which were delineated on the basis of field counts obtained whenever a new group entered the sample. Within groups, generally five to six clusters, with three to four dwellings per cluster, were selected.

The rural cluster stage was identified early in the redesign program as a possible candidate for elimination, on the grounds that (i) the sample variance would be reduced due to having one less stage of sampling, and (ii) the lead time required to introduce new groups into the sample could be shortened from 13 months to 7 months to facilitate using the LFS frame for other household surveys.

A field study was carried out on a sample of groups entering the LFS sample. The study assessed the feasibility of maintaining good quality dwelling lists for entire rural EAs, and examined costs under such a procedure, with positive results on both counts. The variance implications of eliminating the cluster stage were also studied. Using 1971 Census data to simulate both the old and alternative design, components of variance were obtained for the Horvitz-Thompson estimator without ratio estimation. The percent reduction in total NSR variance under the alternative design was found to range from 20%-25% for major labour force characteristics (Choudhry, Lee and Drew, 1985).

Les essais sur le terrain ont eu lieu pendant la période 1982-1984 auprès d'un petit nombre de ménages qui faisaient partie de l'échantillon de l'EPA à ce moment-là. Comme le principal objectif de l'essai était d'évaluer les conséquences que le changement de technique aurait sur la qualité des données, on a divisé la tâche des intervieweurs en deux, une moitié des interviews devant être faites par téléphone et l'autre moitié, sur place. Il est ressorti de l'essai que les taux de non-réponse étaient légèrement inférieurs pour l'échantillon interviewé par téléphone et qu'il n'y avait pas de différences perceptibles dans les estimations relatives aux caractéristiques de la population active.

La décision d'étendre les interviews téléphoniques a eu deux conséquences : une augmentation de la tâche de l'ordre de 10 à 15%, portant cette dernière à 55 à 65 logements en moyenne, et l'attribution de numéros de renouvellement au premier degré d'échantillonnage dans les régions rurales pour que tous les logements dans les unités secondaires reçoivent le même numéro, de manière à réduire le nombre de visites dans ces logements pendant les mois 2 à 6 du cycle où ils faisaient partie de l'échantillon.

4.2.2 Élimination d'un degré d'échantillonnage dans les régions rurales

Dans l'ancien plan de sondage, il y avait trois degrés d'échantillonnage pour les régions rurales des UPE : on prélevait d'abord les unités secondaires (secteurs de dénombrement du recensement), puis les grappes, et enfin les logements. Les grappes correspondaient à des superficies identifiables comptant jusqu'à 20 logements, et leurs limites étaient établies en fonction des résultats du dénombrement sur le terrain effectué chaque fois qu'une nouvelle unité secondaire était introduite dans l'échantillon. Dans les unités secondaires, on prélevait généralement 5 ou 6 grappes comprenant 3 ou 4 logements chacune.

On a déterminé dès le début du programme de remaniement que le degré d'échantillonnage des grappes dans les régions rurales pourrait être éliminé pour les raisons suivantes : (i) la variance de l'échantillon diminuerait du fait qu'il y aurait un degré d'échantillonnage de moins et (ii) le préavis requis avant l'introduction des nouvelles unités secondaires dans l'échantillon pourrait être réduit de 13 à 7 mois de manière à faciliter l'utilisation de la base de sondage de l'EPA aux fins d'autres enquêtes- ménages.

Une étude sur le terrain a été menée auprès d'un échantillon d'unités secondaires qui entraient dans l'échantillon de l'EPA en vue de déterminer s'il était possible de tenir à jour des listes de logements de bonne qualité pour des SD ruraux entiers et combien il en coûterait de le faire. Les résultats ont été positifs aux deux points de vue. On a également étudié l'effet sur la variance de l'élimination du degré d'échantillonnage des grappes. En se servant des données du recensement de 1971 pour simuler à la fois l'ancien plan de sondage et celui qu'on envisageait, on a obtenu les composantes de la variance pour l'estimateur d'Horvitz-Thompson sans estimation par quotient. On a trouvé que la réduction en pourcentage de la variance totale dans les secteurs NAR si l'on appliquait le plan de sondage à l'étude serait de 20 à 25% pour les principales caractéristiques de la population active (Choudhry, Lee et Drew 1985).

4.2.3 Design with urban/rural stratification

The design during the 1970s was the mixed design described earlier. Under this design both strata and PSUs were mixtures of urban and rural areas, with the attempt to maintain the stratum proportion of rural and urban for individual PSUs. A premise underlying this design was that the PSU should correspond to an interviewer's assignment. However, in practice this correspondence is weakened since, in order to attain the desired urban/rural ratio, the urban and rural portions of PSUs cannot always be made contiguous.

As an alternative to this mixed type design (with the rural cluster stage eliminated), the design featuring explicit rural/urban stratification was studied. Under this design, PSUs in both rural and urban strata are designed to yield full interviewer assignments. Although this design features more clustering, particularly of the urban sample, it was suspected that any increase in variance would be offset by reduction in costs due to the more compact PSUs.

Variances under the two designs were studied using Census data for simulation purposes. Also cost implications of the new design were studied. This was achieved using results of a detailed Time and Cost Study and by simulating the dispersion of the sample under the two designs up to the second stage of sampling using population centroids of Enumeration Areas.

A priori expectations were confirmed, and the urban/rural design was found to outperform the old design from a combined cost-variance perspective (Choudhry, Lee and Drew, 1985).

4.2.4 Special design for Prince Edward Island

For Canada's smallest province, Prince Edward Island, sampling rates of 4% are required to produce monthly labour force estimates with required levels of data reliability. In view of the high sampling rates, a less clustered design consisting of a two-stage sample of EAs and dwellings with a deep geographic stratification was adopted. EAs were chosen as stratification units, without regard to rural/urban distinctions, since the urban centres were small and were similar in characteristics to adjacent rural EAs. A total of 12 strata were formed. From each stratum, six EAs were chosen with an average sample yield of 17 dwellings per EA.

4.3 Selection of Primary Sampling Units

As described earlier, the composition of NSR Primary Sampling Units differed for the three principal types of strata. In rural strata, PSUs were collections of geographically compact EAs, in urban strata they were individual or combined urban centres, and in

4.2.3 Plan avec stratification urbaine/rurale

Le plan des années 70 était, comme nous l'avons vu, un plan mixte selon lequel on trouvait tant dans les strates que dans les UPE à la fois des régions urbaines et des régions rurales, et on essayait de faire en sorte que le ratio de la population rurale à la population urbaine dans chaque UPE soit le même que celui dans la strate. Un des buts recherchés était qu'une UPE corresponde à la tâche d'un intervieweur. Toutefois, en pratique, cette correspondance ne pouvait pas toujours être obtenue, car, pour avoir le ratio urbain-rural voulu, il fallait souvent former des UPE dont la partie urbaine et la partie rurale n'étaient pas contiguës.

On a étudié comme solution de rechange à ce plan mixte un plan de sondage avec stratification rurale urbaine explicite (et élimination du degré d'échantillonnage des grappes dans les régions rurales). Selon ce plan, les UPE sont conçues de manière à donner un échantillon correspondant à la tâche complète d'un intervieweur à la fois dans les strates rurales et dans les strates urbaines. Bien que cela implique une augmentation de la taille des UPE, particulièrement dans l'échantillon urbain, on soupçonnait que toute augmentation de la variance serait compensée par une diminution des coûts du fait que les UPE seraient plus compactes.

On a comparé les variances obtenues selon les deux plans de sondage en simulant ces derniers à l'aide des données de recensement. On a aussi étudié les coûts qu'entraînerait le nouveau plan de sondage d'après les résultats d'une analyse temps-coûts détaillée et en simulant la dispersion de l'échantillon dans les deux plans de sondage jusqu'au deuxième degré d'échantillonnage au moyen des coordonnées des endroits, dans les secteurs de dénombrement, où la population était élevée.

Les hypothèses posées a priori ont été confirmées et le plan urbain rural s'est avéré plus performant que l'ancien plan tant du point de vue des coûts que du point de vue de la variance (Choudhry, Lee et Drew 1985).

4.2.4 Plan special pour l'Île-du-Prince-Édouard

Pour la plus petite province du Canada, l'Île-du-Prince-Édouard, il faut avoir des taux de sondage de 4% pour produire des estimations mensuelles de l'EPA atteignant le niveau de fiabilité voulu. Comme il faut des taux de sondage si élevés, on a adopté un plan prévoyant un plus grand nombre de petites grappes qui consistait en un échantillon à deux degrés de SD et de logements et supposait une stratification géographique détaillée. On a pris les SD comme unités de stratification sans se préoccuper de la distinction rurale urbaine, car les centres urbains étaient petits et avaient des caractéristiques semblables à celles des SD ruraux voisins. Douze strates ont été formées en tout. Six SD ont été prélevés dans chaque strate et chaque SD avait en moyenne un rendement de l'échantillon de 17 logements.

4.3 Sélection des unités primaires d'échantillonnage

Comme nous l'avons vu dans le chapitre sur la stratification, la composition des unités primaires d'échantillonnage dans les secteurs NAR variait dans chacune des trois principales sortes de strates. Dans les strates rurales, les UPE étaient des ensembles de SD compacts du point de vue

mixed strata they were collections of geographically compact rural EAs, together with one or more nearby urban centres or portions thereof.

Regardless of the stratum type, the methodology for selection of PSUs was the same. As was the case in the previous LFS design, the Randomized Probability Proportional to Size Systematic (RPPSS) method due to Hartley and Rao (1962) was used. Principal reasons for the choice of the method are its flexibility to accommodate rotation of sampling units, and to accommodate changes in the sample size through increase or decrease in the number of sampled units.

In each stratum, either two or three PSUs were selected, whichever resulted in an expected sample yield per PSU closer to the desired 55-65 dwellings. Determination of the number of PSUs to be selected per stratum and the actual sampling procedure is as follows. Denoting

R_h = the within stratum inverse sampling rate

D_h = the number of dwellings in stratum h

D_{ih} = the number of dwellings in PSU stratum h

the number of PSUs to select n_h (= 2 or 3) was determined as

$$n_h = \frac{D_h}{d_h R_h}$$

where d_h is the expected sample size and is as close as possible to 55-65 dwellings.

Selection of the n_h PSUs then proceeded as follows:

- (i) Within each stratum the list of PSUs is randomized.
- (ii) The inverse sampling ratio for each PSU is calculated as

$$R_{ih} = \left[\frac{R_h n_h D_{ih}}{D_h} \right]$$

or

$$\left[\frac{R_h n_h D_{ih}}{D_h} \right] + 1$$

so that

$$\sum_i R_{ih} = n_h R_h$$

(Here $[x]$ refers to the greatest integer part of x .)

- (iii) Using R_{ih} as the size measure, R_h as the sampling interval, and by choosing a random integer r_i ($1 \leq r_i \leq R_h$), a pps systematic sample of n_h PSUs is selected from each stratum.

géographique, dans les strates urbaines, elles étaient constituées d'un ou de plusieurs centres urbains et, dans les strates mixtes, elles étaient formées de quelques SD ruraux compacts sur le plan géographique et d'un ou de plusieurs centres urbains adjacents, certains entiers et d'autres non.

La méthode utilisée pour le prélèvement des UPE a été la même quel que soit le type de strate. Comme dans l'ancien plan de sondage de l'EPA, on a pris la méthode d'échantillonnage systématique avec classement aléatoire et probabilité proportionnelle à la taille (ESCAPPT) que l'on doit à Hartley et Rao (1962). La principale raison qui a motivé ce choix est la souplesse de cette méthode quand il faut procéder au renouvellement des unités d'échantillonnage ou modifier la taille de l'échantillon en augmentant ou en diminuant le nombre d'unités échantillonnées.

Dans chaque strate, on a choisi 2 ou 3 UPE, selon ce qui permettait d'obtenir un rendement prévu de l'échantillon par UPE qui s'approchait le plus des 55 à 65 logements voulus. On a déterminé le nombre d'UPE à prélever dans chaque strate et procédé à l'échantillonnage de la manière suivante. Soit

R_h = la fraction de sondage inverse à l'intérieur de la strate

D_r = le nombre de logements dans la strate h

D_{ih} = le nombre de logements dans l'UPE i , strate h

On a déterminé le nombre d'UPE à prélever n_h (= 2 ou 3) par la formule

$$n_h = \frac{D_h}{d_h R_h}$$

où d_h est la taille prévue de l'échantillon et se rapproche le plus possible de 55 à 65 logements.

Pour prélever les n_h UPE, on a procédé de la manière suivante :

- (i) A l'intérieur de chaque strate, on a randomisé la liste des UPE.
- (ii) On a calculé la fraction de sondage inverse pour chaque UPE par la formule

$$R_{ih} = \left[\frac{R_h n_h D_{ih}}{D_h} \right]$$

ou

$$\left[\frac{R_h n_h D_{ih}}{D_h} \right] + 1$$

de manière à ce que

$$\sum_i R_{ih} = n_h R_h$$

Ici, $[x]$ représente la partie entière de x .)

- (iii) En prenant les R_{ih} comme mesures de la taille et R_h comme pas de sondage et en choisissant un nombre entier aléatoire r_i ($1 \leq r_i \leq R_h$), on a prélevé dans chaque strate un échantillon systématique avec ppt de n_h UPE.

4.4 Sub-Sampling in Rural Strata

4.4.1 Sampling within Primary Sampling Units

The secondary sampling unit (SSU) in rural strata is termed the group. A group generally corresponds to a 1981 Census Enumeration Area, with some collapsing in case of EAs with low dwelling counts, and with splitting of EAs with high counts. The rules governing combining and splitting of EAs were that EAs with less than 20 dwellings were collapsed with a neighboring EA, while EAs whose dwelling count was more than one-third of the dwelling count for the PSU were split into parts having similar dwelling counts. The splitting was done using recognizable physical features on topographical maps, with the number of dwellings per split portion being estimated on the basis of the dwellings plotted on the topographical maps.

As described earlier, average populations of PSUs differed substantially depending on the province and to a lesser extent on the Economic Region within province. The different PSU populations led to different within-PSU sampling rates in order to achieve the desired sample yields of 50-60 dwellings. Further, it was desired to select between 10-20 dwellings per group, a design parameter that was left unchanged from the previous design, and was meant to represent the workload an interviewer could be expected to handle in one day's work. The fixed yields per PSU and per SSU combined with different strata and PSU sizes across the country, led to dramatically different second stage sampling rates ranging from a sub-sampling rate of close to 1 in the Atlantic Provinces (excluding Prince Edward Island where a special design was adopted), to sampling rates of close to 1.3 in Quebec and Ontario.

Selection of groups followed the same method used for PSUs, namely, the randomized PPS systematic method. The first step was to randomly order the rural groups for the selected PSU.

Next, n_{hi} , the number of groups to select per PSU was determined as follows:

- (i) If N_{hi} , the number of groups in PSU_{hi} is 5 or less, then select all of them.
- (ii) If the PSU_{hi} has 6 or more groups, select 3 groups if this results in an expected sample yield of 20 or less, otherwise select 6 groups. The rationale for selection of 3 or 6 groups relates to assignment of rotation numbers as discussed in the next chapter.
- (iii) If the proportion of groups selected via rule (ii) is greater than or equal to two-thirds, then select all groups.

4.4 Sous-échantillonnage dans les strates rurales

4.4.1 Échantillonnage dans les UPE

L'unité secondaire d'échantillonnage (USE) dans les strates rurales s'appelle le groupe. Un groupe correspond généralement à un secteur de dénombrement du recensement de 1981, mais il peut s'agir de plusieurs SD dans lesquels le nombre de logements était faible et que l'on a réunis ou encore d'une partie d'un SD dans lequel le nombre de logements était élevé. Les critères sur lesquels on s'est basé pour déterminer s'il fallait grouper ou diviser les SD étaient les suivants : les SD dans lesquels il y avait moins de 20 logements ont été rattachés à un SD adjacent et les SD dans lesquels le nombre de logements était supérieur au tiers du nombre de logements dans l'UPE ont été divisés en parties plus ou moins égales du point de vue du nombre de logements. La division a été faite en fonction des accidents de terrain reconnaissables sur les cartes topographiques, le nombre de logements par partie formée étant estimé d'après les indications à cet effet figurant sur ces cartes.

Comme nous l'avons déjà vu, la population moyenne des UPE variait nettement selon la province et, à l'intérieur des provinces, quoique dans une moindre mesure, selon la région économique. À cause de cette variation, il a fallu prendre des fractions de sondage différentes dans les UPE pour obtenir le rendement de l'échantillon voulu, soit 50 à 60 logements. En outre, on voulait prélever de 10 à 20 logements par groupe, comme dans l'ancien plan de sondage, car on estimait que c'était là la charge de travail qu'un intervieweur pouvait accomplir en une journée. Le fait que le rendement de l'échantillon dans les UPE et les USE était fixe et que la taille des strates et des UPE était différente d'un bout à l'autre du pays a entraîné une fluctuation marquée des fractions de sondage au deuxième degré d'échantillonnage, lesquelles ont varié de près de 1 dans les provinces de l'Atlantique (exception faite de l'Île-du-Prince-Édouard où on a employé un plan spécial) à près de 1.3 au Québec et en Ontario.

On a appliqué la même méthode pour prélever les USE ou groupes que pour les UPE, soit la méthode d'échantillonnage systématique avec classement aléatoire et ppt. La première étape a consisté à mettre dans un ordre aléatoire les groupes ruraux dans les UPE choisies.

On a ensuite déterminé n_{hi} , le nombre de groupes à prélever par UPE, de la manière suivante :

- (i) Si N_{hi} , le nombre de groupes dans une UPE_{hi} , est égal ou inférieur à 5, prélever tous les groupes.
- (ii) Si l' UPE_{hi} compte 6 groupes ou plus, prélever trois groupes si cela permet d'obtenir un rendement prévu de l'échantillon de 20 logements ou moins; autrement, prélever 6 groupes. La raison pour laquelle on prend 3 ou 6 groupes est liée à l'attribution des numéros de renouvellement, dont il est question dans le prochain chapitre.
- (iii) Si la proportion de groupes prélevés par la règle (ii) est égale ou supérieure aux deux tiers, prélever tous les groupes.

When all groups in a PSU are selected, the PSU inverse sampling ratio R_{hi} is adopted as the inverse sampling ratio for each group. Where only a fraction of the groups are selected, the inverse sampling ratio R_{hij} is determined as the greatest integer part of

$$\frac{n_{hi} D_{hij} R_{hi}}{D_{hi}}$$

or

$$\frac{n_{hi} D_{hij} R_{hi}}{D_{hi}} + 1,$$

chosen so that

$$\sum_j R_{hij} = n_{hi} R_{hi}$$

4.4.2 Selection of dwellings

The procedure followed for the selection of dwellings within the selected groups in the NSR areas is same as that used within the clusters in the SR areas, that is, dwellings are selected systematically by applying a random start and a sampling interval equal to the inverse sampling ratio of the selected group.

4.5 Sub-Sampling in Urban Strata

4.5.1 Clustering of urban centres

Primary sampling units in urban strata comprise generally one, but sometimes two or three urban centres. Within selected urban PSUs, each urban centre is clustered, and sampled separately at the PSU sampling rate.

At the time of the sample redesign, procedures for clustering of NSR urban areas were the same as those followed for the smaller SRUs lying outside the block face coverage; that is, for the most part, 1981 Census visitation records were used to establish dwelling counts, with field counting being resorted to whenever the visitation records or accompanying maps lacked sufficient detail.

NSR urban clusters in general contained fewer dwellings than those in SR areas. Minimum cluster sizes were set at three dwellings, with maximums of three times the PSU inverse sampling ratio.

4.5.2 Selection of clusters

Clusters are selected independently from each urban centre belonging to the sampled PSUs using the randomized pps systematic method. Steps in cluster selection are as follows:

- (i) Within each selected urban centre, the list of clusters is randomized.

Lorsque tous les groupes dans une UPE sont choisis, la fraction de sondage inverse de l'UPE R_{hi} devient la fraction de sondage inverse de chaque groupe. Lorsqu'une partie seulement des groupes est choisie, on détermine les fractions de sondage R_{hij} en prenant la partie entière de

$$\frac{n_{hi} D_{hij} R_{hi}}{D_{hi}}$$

ou

$$\frac{n_{hi} D_{hij} R_{hi}}{D_{hi}} + 1,$$

de manière à ce que

$$\sum_j R_{hij} = n_{hi} R_{hi}$$

4.4.2 Selection des logements

Pour prélever les logements dans les groupes choisis dans les secteurs NAR, on a procédé de la même manière que pour les logements dans les grappes dans les secteurs AR, c'est-à-dire qu'on a choisi les logements systématiquement en prenant une origine aléatoire et un pas de sondage égal à la fraction de sondage inverse des groupes choisis.

4.5 Sous-échantillonnage dans les strates urbaines

4.5.1 Formation des grappes dans les centres urbains

Les unités primaires d'échantillonnage sont généralement formées d'un centre urbain, mais il arrive qu'elles en comprennent deux ou trois. Dans les UPE urbaines choisies, on forme les grappes et on prélève l'échantillon séparément dans chaque centre urbain en appliquant la fraction de sondage de l'UPE.

Au moment du remaniement du plan de sondage, la méthode utilisée pour former les grappes dans les centres urbains des secteurs NAR était la même que celle dont on se servait pour les petites UAR où l'unité de sondage n'était pas le côté d'îlot. Autrement dit, dans la plupart des cas, on s'est servi des registres des visites du recensement de 1981 pour établir les chiffres sur les logements et on a eu recours au dénombrement sur le terrain lorsque les registres des visites ou les cartes qui les accompagnaient n'étaient pas suffisamment détaillées.

En général, les grappes dans les centres urbains des secteurs NAR contenaient moins de logements que celles dans les secteurs AR. La taille minimum d'une grappe a été fixée à 3 logements et la taille maximum, à 3 fois la fraction de sondage inverse de l'UPE.

4.5.2 Sélection des grappes

On a prélevé les grappes de façon indépendante dans chaque centre urbain faisant partie d'une UPE échantillonnée par la méthode d'échantillonnage systématique avec classement aléatoire et ppt. La sélection des grappes s'est faite selon les étapes suivantes :

- (i) Dans chaque centre urbain choisi, on randomise la liste des grappes.

- (ii) The desired expected sample yield (or density factor, d) is taken as three dwellings per cluster. The number of clusters to select in PSU_i , stratum h , urban centre j is determined as

$$d_{hij} = \frac{D_{hij}}{n_{hij} R_{hi}}$$

where the density factor is as close as possible to 3.

- (iii) Letting k denote a cluster, all clusters for which the number of dwellings D_{hijk} was greater than or equal to $\Sigma_k D_{hijk}/n_{hij}$ were selected with certainty, and were assigned an inverse sampling ratio equal to that for the PSU.

- (iv) For the remaining clusters an inverse sampling ratio, R_{hijk} , was obtained as the greatest integer part of

$$\frac{D_{hijk} R_{hij}}{d^*_{hij}}$$

or

$$\frac{D_{hijk} R_{hij}}{d^*_{hij}} + 1,$$

where d^*_{hij} was defined analogously to d_{hij} , but only for those n_{hij} clusters not selected with certainty. The individual R_{hijk} 's calculated at this step were chosen to total $n_{hij} R_{hij}$.

- (v) Using R_{hijk} as the cluster size, a systematic sample of clusters was selected in the usual manner.

4.6 Sub-Sampling in Mixed Strata

In mixed strata, rural and urban portions of the selected PSUs are sampled separately. This, together with the design constraint that all PSUs within a stratum have roughly equal proportions of rural and urban population, amounts to an implicit urban-rural stratification.

The sampling of rural portions of the PSUs proceeds in a very similar fashion to the sampling of PSUs in rural strata, with the exception that two or four selected groups are generally aimed for (instead of three or six), since the rural portions of mixed PSUs have correspondingly fewer dwellings.

The sampling of urban portions of PSUs also proceeds similarly to the sampling of urban PSUs, although special steps are necessary for treatment of urban centres which, when designing PSUs, were fractionally associated with two or more PSUs.

- (ii) On prend comme rendement prévu de l'échantillon voulu (ou facteur de densité d) trois logements par grappe. On détermine le nombre de grappes à choisir dans l'UPE, strate h , centre urbain j par la formule

$$d_{hij} = \frac{D_{hij}}{n_{hij} R_{hi}}$$

où le facteur de densité se rapproche autant que possible de 3.

- (iii) Soit k étant la grappe, on choisit avec une probabilité égale à 1 toutes les grappes où le nombre de logements D_{hijk} est supérieur ou égal $\Sigma_k D_{hijk}/n_{hij}$ et on leur attribue une fraction de sondage inverse égale à celle de l'UPE.

- (iv) On attribue aux grappes restantes la fraction de sondage inverse R_{hijk} qu'on obtient en prenant la partie entière de

$$\frac{D_{hijk} R_{hij}}{d^*_{hij}}$$

ou

$$\frac{D_{hijk} R_{hij}}{d^*_{hij}} + 1,$$

où l'on définit d^*_{hij} de façon analogue à d_{hij} , mais seulement pour les n_{hij} grappes non choisies avec une probabilité égale à 1. Les R_{hijk} calculées à cette étape-ci sont corrigées de nouveau pour que leur somme soit égale à $n_{hij} R_{hij}$.

- (v) En prenant R_{hijk} comme taille des grappes, on prélève un échantillon systématique de grappes de la manière habituelle.

4.6 Sous-échantillonnage dans les strates mixtes

Dans les strates mixtes, les parties rurales et urbaines des UPE choisies sont échantillonnées séparément. Cette façon de procéder, ainsi que la contrainte établie dans le plan de sondage selon laquelle le ratio de la population rurale à la population urbaine soit à peu près égal dans toutes les UPE d'une strate, équivaut à une stratification urbaine rurale implicite.

L'échantillonnage des parties rurales des UPE se fait de façon très semblable à l'échantillonnage des UPE dans les strates rurales, sauf qu'on cherche à prélever 2 ou 4 groupes (plutôt que 3 ou 6) du fait qu'il y a moins de logements dans la partie rurale des UPE mixtes.

L'échantillonnage des parties urbaines des UPE se fait aussi de façon semblable à l'échantillonnage des UPE dans les strates urbaines, sauf qu'il faut prendre des mesures spéciales à l'égard des centres urbains qui ont été divisés au moment de la formation des UPE et associés à deux UPE ou plus.

In such cases, the entire urban centre is clustered in the usual fashion. Say the urban centre is divided into m parts at the PSU delineation stage, which are assigned to m^* PSUs (with $m^* \leq m$, since some PSUs may have received more than one part). Then the clusters are randomly assigned to each one of these parts. The partitioning was done to equalize the number of dwellings, as opposed to the number of clusters. Each PSU sharing the urban centre was then assigned one or more parts as required.

The clusters assigned to the selected PSU in this fashion were sampled at the PSU inverse sampling rate using the same procedures as described earlier for urban strata.

4.7 Sample Design for Special Areas

Remote areas of most provinces, and institutions, such as hospitals, schools and military establishments were designated as Special Areas. For sampling purposes, the special areas are divided into two parts: (1) Remote Areas, and (2) Institutions, which are sampled independently.

Remote areas, consisting of the northernmost parts of all provinces except Prince Edward Island, Nova Scotia and New Brunswick, cover a vast land area but contain less than 1% of the population. For military establishments and institutions only the civilian non-inmate population is included in the survey.

4.7.1 Rationale for special areas

Two primary reasons for sampling the Special Areas separately from the regular (SR and NSR) areas were:

- (i) The population living in these areas possesses characteristics differing from those of the rest of the population and therefore, if it were included in the regular areas, the sampling variance would increase. Further, if not segregated, the special areas would introduce instability into the sample estimates, especially at the time when units rotate in or out.
- (ii) There have been instances when, due to inaccessibility or weather, the method and/or the frequency of data collection has had to be modified. For example, interviews may be collected only once in a six-month cycle, with the data being carried forward during intervening months.

4.7.2 Design of the remote areas

Each of the most northerly Economic Regions of those provinces with remote areas were examined individually. Cities with unique characteristics (such as Fort McMurray, Flin Flon, The Pas, Thompson and

Dans ces cas-là, le centre urbain entier est réparti en grappes de la manière habituelle. Supposons que le centre urbain a été divisé en m parties à l'étape de la délimitation des UPE et que celles-ci ont été attribuées à m^* UPE (où $m^* \leq m$, car il se peut qu'ont ait attribué plus d'une partie à certaines UPE). On a ensuite attribué les grappes au hasard à chacune des m parties. Le but de la partition était d'égaliser le nombre de logements, par opposition au nombre de grappes par parties. On a ensuite attribué à chacune des UPE qui se partageaient le centre urbain une ou deux parties selon le besoin.

On a échantillonné les grappes attribuées de cette manière aux UPE choisies en appliquant la fraction de sondage inverse de l'UPE et en utilisant la même méthode que celle que nous avons décrite dans la section sur les strates urbaines.

4.7 Plan de sondage pour les secteurs spéciaux

On a décidé que les régions éloignées de la plupart des provinces et les établissements institutionnels comme les hôpitaux, les maisons d'enseignement et les établissements militaires constituaient des secteurs spéciaux. Pour les besoins de l'échantillonnage, on a divisé les secteurs spéciaux en deux catégories, les régions éloignées et les établissements institutionnels, qu'on a échantillonnées séparément.

Les régions éloignées, qui représentent la partie la plus au Nord de toutes les provinces à l'exception de l'Île-du-Prince-Édouard, de la Nouvelle-Écosse et du Nouveau-Brunswick, ont une vaste superficie mais comptent moins de 1% de la population. En ce qui concerne la population des établissements militaires et des établissements institutionnels, seuls les civils et les personnes autres que les pensionnaires de ces établissements ont été inclus dans l'enquête.

4.7.1 Raison d'être des secteurs spéciaux

Les deux principales raisons pour lesquelles les secteurs spéciaux ont été échantillonnés séparément des secteurs normaux (AR et NAR) étaient les suivantes :

- (i) La population de ces secteurs possède des caractéristiques différentes de celles de la population en général. Par conséquent, si on avait inclus ces secteurs avec les autres, la variance d'échantillonnage aurait été plus grande. Cela aurait aussi introduit un élément d'instabilité dans les estimations d'échantillon, surtout au moment où les unités sont introduites dans l'échantillon ou en sont retirées par renouvellement.
- (ii) Il est déjà arrivé qu'on doive modifier la méthode de collecte des données ou la fréquence de cette collecte à cause du mauvais temps ou de l'inaccessibilité de ces secteurs. Ainsi, dans certains cas, les interviews n'ont lieu qu'une fois dans le cycle de 6 mois et les données sont reportées pour les autres mois.

4.7.2 Plan de sondage pour les régions éloignées

On a examiné individuellement chacune des RE situées le plus au Nord des provinces où il y a des régions dites éloignées. On a échantillonné par la méthode normale qui s'appliquait au type de secteur en cause les villes qui avaient

Happy Valley), an urban stratum in Quebec and one rural stratum in each of British Columbia, Alberta, Ontario and Quebec, were subjected to the regular sampling procedure for the corresponding type of area.

EAs in the remote areas of most provinces tend to be very large geographic areas, with the population clustered in small settlements or "places". Enumeration of the population outside these places is extremely expensive and often not possible due to the remoteness. It was decided, therefore, to consider places as units of selection, with the precise boundaries to be determined by the local interviewer at the time of selection.

The units of selection for Remote Areas in all provinces except Ontario and Alberta consist of all places with a population of 25 or more. In Ontario and Alberta it was found that a frame using places would exclude too much population (almost 50% in these provinces) and there it was decided to use EAs or EA combinations with 20 or more occupied dwellings as the unit of selection. A similar problem existed in British Columbia; however, by including five very large EAs in the frame instead of only the places within them, the coverage of remote areas in this province exceeded 90%.

The resulting design reduces the difficulties of enumerating the very dispersed population while still ensuring more than 96% coverage in the Remote Areas, which is certainly adequate, given that the Remote Areas themselves represent only about 1% of the population of Canada.

The measure of size for Remote Areas is the number of households (occupied dwellings from the 1981 Census). Units (place or EA combination) are listed in descending order of size, and two, four or six units are selected in order to ensure an expected sample take of between 10 and 15 households from each selection. The larger units may be sub-divided into clusters by splitting EAs. Only two or more clusters are then selected, thus reducing the travel and maintenance costs of listing.

4.7.3 Design of the institutions frame

Certain EAs, consisting of one or more collective dwellings of substantial size, such as hospitals, military establishments (excluding permanent married quarters) and other institutions, were designated as Collective EAs in the 1981 Census. From all Collective EAs only those which contain military personnel or inmates (such as jails, hospitals, sanatoria etc.) were considered as part of the Institutions Frame in the special areas. Other collective EAs were deemed to contain "regular" population and were sampled with the regular SR and NSR sample.

des caractéristiques uniques en leur genre (comme Fort McMurray, Flin Flon, The Pas, Thompson et Happy Valley), une strate urbaine au Québec et une strate rurale en Colombie-Britannique, en Alberta, en Ontario et au Québec.

En général, les SD dans les régions éloignées de la plupart des provinces sont très étendus et leur population est regroupée dans de petites agglomérations ou "localités". Le dénombrement de la population qui habite en dehors de ces localités est une opération qui coûte extrêmement cher et qui est souvent impossible à cause de l'éloignement de cette population. Il a donc été décidé de considérer ces localités comme des unités de sélection dont les limites exactes seraient déterminées par l'intervieweur sur place, au moment où elles seraient choisies.

Les unités de sélection dans les régions éloignées de l'ensemble des provinces sauf l'Ontario et l'Alberta sont les localités dont la population est d'au moins 25 habitants. En Ontario et en Alberta, on a trouvé que, si on constituait une base dans laquelle figureraient seulement les localités, une trop forte proportion de la population serait exclue (près de 50%); on a donc décidé de prendre comme unités de sélection les SD ou des groupes de SD comptant 50 logements occupés ou plus. On a rencontré un problème semblable en Colombie-Britannique, et on s'est rendu compte que, en incluant dans la base de sondage cinq très grands SD au lieu d'y mettre seulement les localités, on obtenait un taux de couverture supérieur à 90% dans les régions éloignées de cette province.

Le plan ainsi mis au point réduit la difficulté de dénombrer les populations très dispersées tout en garantissant un taux de couverture de plus de 96% dans les régions éloignées, ce qui est certainement adéquat compte tenu du fait qu'on ne trouve dans ces régions que 1% de la population du Canada.

La taille des unités dans les régions éloignées est déterminée par le nombre de ménages (logements occupés selon les résultats du recensement de 1981). Les unités (localités ou combinaisons de SD) sont énumérées par ordre décroissant de taille, et deux, quatre ou six unités sont prélevées pour assurer un rendement prévu de l'échantillon de 10 à 15 ménages par unité. Les grandes unités peuvent être subdivisées en grappes en fractionnant les SD. On prélève deux grappes ou plus, ce qui réduit les frais de déplacement et de mise à jour des listes.

4.7.3 Conception de la base d'établissements institutionnels

Au recensement de 1981, il avait été décidé que certains SD, qui consistaient en un ou plusieurs logements collectifs de grande taille, comme les hôpitaux, les établissements militaires (exception faite des logements familiaux sur les bases militaires) et d'autres établissements institutionnels, portaient le nom de SD collectifs. De tous les SD collectifs, seuls ceux où se trouvaient des militaires ou des pensionnaires d'établissement (p. ex. les prisons, les hôpitaux et les sanatoria) ont été considérés comme faisant partie de la base d'établissements institutionnels dans les secteurs spéciaux. Pour ce qui est des autres, on a jugé qu'ils contenaient une population "normale" et qu'ils devraient être échantillonnés avec le reste de l'échantillon des secteurs AR et NAR.

A provincial level list, in descending order of size, was created for the Institutions frame in all provinces except Prince Edward Island. In PEI, due to insufficient population, all collective EAs are treated as components of the regular SR and NSR frames. The size measure used was the labour force population 15 years of age and older, which excludes all military, institutional and overseas population. Collective EAs with labour force population less than two (less than four in Ontario) were omitted from the sample frame of institutions, and as such stood no chance of selection. As is the case with remote areas, dropping the very small units was inconsequential in terms of the overall survey coverage. Two or more units per stratum were selected using PPS sampling, with an expected sample take of between 10 and 20 persons where possible.

A summary of Special Areas is found in Appendix E.

On a dressé une liste par ordre décroissant de taille des SD collectifs dans la base d'établissements institutionnels pour chacune des provinces sauf l'Île-du-Prince-Édouard. À l'Île-du-Prince-Édouard, l'effectif de la population étant trop faible, on a traité tous les SD collectifs comme s'ils faisaient partie de la base normale des secteurs AR et NAR. Pour mesurer la taille des secteurs, on a pris la population active âgée de 15 ans et plus moins le personnel militaire, les pensionnaires des établissements institutionnels et les personnes à l'étranger. Les SD collectifs dont la population active était inférieure à 2 (ou à 4 en Ontario) ont été exclus de la base de sondage des établissements institutionnels, de sorte qu'ils n'avaient aucune chance d'être choisis. Comme dans le cas des régions éloignées, le fait de laisser tomber les toutes petites unités a été sans conséquence sur le taux de couverture globale de l'enquête. On a prélevé deux unités ou plus par strate selon une probabilité proportionnelle à la taille en visant un rendement prévu de l'échantillon de 10 à 20 personnes là où c'était possible.

Le traitement des secteurs spéciaux est résumé à l'annexe E.

Chapter 5

Sample Size Determination

5.1 Introduction

As part of the revision of the Labour Force Survey during the 1970s, the sample size was expanded from 36,400 to 55,700 households per month. The principal objective of the sample expansion, which was carried out from 1976 to 1977, was to improve the reliability of provincial estimates. Prior to the sample size expansion, the Labour Force Survey (LFS) sample had been allocated with the objective of providing good national estimates, although sufficient oversampling of smaller provinces was done to permit the release of publishable data for all provinces.

The strategy adopted in expanding the sample size was to keep the allocation of the initial 36,400 households intact, and to allocate an additional 19,300 households so as to meet or exceed a minimum level of data reliability in all provinces. Analysis showed that a target of an 8% or better coefficient of variation (CV) for monthly estimates of unemployed was attainable in all provinces except Prince Edward Island; this province would have required sampling rates in excess of what was felt tolerable in terms of response burden on the population. Allocation of the increased sample in this fashion led to decreases in the CVs for unemployed by 40% in Prince Edward Island and Saskatchewan; 25% to 33% in Newfoundland, Manitoba, and Alberta; 10% in Nova Scotia, New Brunswick and British Columbia; and 5% in Quebec and Ontario.

Within each province, separate sampling rates were adopted for NSR and SR areas, and deviations from these uniform sampling rates occurred only in a few exceptional cases. This was in keeping with the objective of maximizing the improvement in the reliability of provincial data. Nevertheless, as a by-product of the sample size increase it became possible to publish some subprovincial data monthly when the data release criterion of a CV of 33% or less was met. In fact, estimates of basic labour force characteristics (persons employed, unemployed and not in labour force; and corresponding rates) were published for CMAs, and for 53 individual or grouped Economic Regions, subject to suppression of cells for which the rounded estimate of level was less than 4,000. Due to the uniform sampling rates within each province, the reliability of subprovincial data varied greatly depending on the population of the subprovincial area in question.

During the late 1970s and early 1980s the importance attached to subprovincial estimates among users of the survey data grew steadily. Important developments in this regard include the use of

Chapitre 5

Détermination de la taille de l'échantillon

5.1 Introduction

Au moment de la révision de l'enquête sur la population active dans les années 70, la taille de l'échantillon a été portée de 36,400 à 55,700 ménages. Le principal objectif de cette opération, qui a été effectuée entre 1976 et 1977, était de rendre les estimations provinciales plus fiables. Auparavant, la répartition de l'échantillon de l'EPA visait principalement la production de bonnes estimations à l'échelle nationale, quoique le suréchantillonnage des petites provinces permettait de diffuser des données publiables sur toutes les provinces.

La stratégie adoptée lorsqu'on a élargi l'échantillon a été de ne pas toucher à la répartition des 36,400 ménages initiaux et de répartir les 19,300 ménages supplémentaires de manière à atteindre ou à dépasser un degré minimum de fiabilité des données dans l'ensemble des provinces. Il est ressorti des analyses qu'on pourrait ainsi obtenir un coefficient de variation (CV) de 8% ou moins pour les estimations relatives aux chômeurs dans toutes les provinces sauf l'Île-du-Prince-Édouard, où il aurait fallu, pour atteindre ce but, utiliser des fractions de sondage dépassant le seuil de tolérance de la population à l'égard du fardeau de réponse. Cette allocation du nouvel échantillon a entraîné une diminution du CV de l'ordre de 40% pour ce qui est des estimations du chômage à l'Île-du-Prince-Édouard et en Saskatchewan, de 25 à 33% pour les estimations concernant Terre-Neuve, le Manitoba et l'Alberta, de 10% pour celles se rapportant à la Nouvelle-Écosse, au Nouveau-Brunswick et à la Colombie-Britannique et de 5% pour celles ayant trait au Québec et à l'Ontario.

Pour atteindre l'objectif d'augmenter au maximum la fiabilité des données provinciales, on a adopté, dans chaque province, une fraction de sondage pour les secteurs AR et une autre pour les secteurs NAR, et on ne s'est écarté de ce taux que dans quelques cas exceptionnels. Une conséquence de l'élargissement de l'échantillon a néanmoins été qu'il est devenu possible de publier des données infraprovinciales sur une base mensuelle lorsque le critère selon lequel le CV devait être de 33% ou moins pour la diffusion des données était respecté. De fait, des estimations relatives aux caractéristiques de base de la population active (nombre de personnes occupées, en chômage et inactives et taux correspondants) ont été publiées pour les RMR et pour 53 régions économiques individuelles ou groupées, la seule contrainte étant la suppression des cellules pour lesquelles les estimations arrondies étaient inférieures à 4000. Comme les fractions de sondage étaient uniformes à l'intérieur d'une province, la fiabilité des données infraprovinciales variait beaucoup selon la population des régions en question.

Pendant la fin des années 70 et le début des années 80, l'importance accordée aux estimations infraprovinciales de l'EPA par les utilisateurs des données de l'enquête s'est accrue de façon continue. Ainsi, à Emploi et Immigration

subprovincial LFS estimates by Employment and Immigration Canada in the administration of the Unemployment Insurance Program, and by the Department of Regional and Industrial Expansion in the calculation of their Development Index. In the former case, three-month moving averages of seasonally adjusted unemployment rates for individual or combined Economic Regions are used to establish regionally variable qualifying periods for weeks worked prior to eligibility for Unemployment Insurance benefits. In the latter case, three-year average estimates of employment and unemployment for individual or grouped census divisions are used in the calculation of the Development Index. This index determines the maximum level of funding assistance for which employers wishing to locate in an area may apply.

Prior to the post-1981 Censal redesign of the LFS, a program evaluation of the survey was carried out that involved extensive consultation with users of the survey data. From the evaluation, and from user consultations conducted at the outset of the sample redesign, it became clear that the predominant interest vis-à-vis the quality of the survey estimates was in improved subprovincial data. Most users were of the opinion that provincial data quality was acceptable, and that gains in subprovincial data should not be achieved at the expense of reduced reliability for provincial data.

5.2 Data Reliability Objectives

Following consultations with the provinces and with key users among the federal government departments noted above, a set of data reliability objectives for the redesigned sample was established. These objectives in turn dictated the overall strategy for determining the size and allocation of the redesigned sample.

The following data reliability objectives were set for the redesigned sample:

- (i) For Canada and each of the provinces, no reduction in the reliability for monthly estimates of level and estimates of month-to-month change for total employment and unemployment.
- (ii) For 24 Census Metropolitan Areas as defined by the 1981 Census, monthly estimates for total unemployment with CVs of 20% or less.
- (iii) For 66 Subprovincial Economic Regions agreed to in consultation with the provinces, monthly estimates for total unemployment with CVs of 25% or less.
- (iv) For Census Agglomerations with populations of 60,000 or more in Quebec and Ontario, and with populations of 25,000 or more in remaining provinces, three-month average estimates for total unemployment with CVs of 25% or less.

Canada, on a commencé à utiliser ces estimations pour l'administration du Programme de l'assurance-chômage et à Industrie, Sciences et Technologie Canada, pour le calcul de l'indice de développement. À Emploi et Immigration, on se sert des moyennes mobiles sur trois mois des taux de chômage désaisonnalisés dans les régions économiques prises individuellement ou en groupe pour établir le nombre de semaines qu'il faut avoir travaillé, selon la région, pour être admissible aux prestations d'assurance-chômage. À Industrie, Science et Technologie, on se sert des estimations moyennes sur trois ans de l'emploi et du chômage dans des divisions du recensement prises individuellement ou en groupe pour calculer l'indice de développement. Celui-ci détermine le montant maximum d'aide financière que peut demander un employeur désirant s'installer dans une région donnée.

Avant le remaniement de l'EPA qui a suivi le recensement de 1981, une évaluation du programme de l'enquête a été faite pour laquelle les utilisateurs des données ont été consultés en grand nombre. Il est ressorti clairement de cette évaluation et des consultations auprès des utilisateurs effectuées au début du remaniement du plan de sondage que ce qui intéressait avant tout ces derniers en ce qui a trait à la qualité des estimations de l'enquête, c'était qu'on améliore les données infraprovinciales. La plupart des utilisateurs étaient d'avis que la qualité des données provinciales était acceptable et qu'il ne fallait pas réduire la fiabilité de ces dernières pour améliorer les données infraprovinciales.

5.2 Objectifs de fiabilité des données

À la suite des consultations mentionnées plus haut et menées auprès des provinces et des principaux ministères fédéraux qui utilisent les données de l'EPA, on a dressé la liste des objectifs de fiabilité des données que le nouveau plan de sondage devait permettre d'atteindre, et ces objectifs ont à leur tour inspiré la stratégie globale adoptée pour déterminer la taille et la répartition de l'échantillon.

Les objectifs de fiabilité des données que le nouveau plan de sondage devait permettre d'atteindre sont les suivants :

- (i) Pour le Canada et les dix provinces, maintenir le degré de fiabilité des estimations mensuelles du niveau de l'emploi et du chômage et des estimations de la variation mensuelle de ce niveau.
- (ii) Pour 24 régions métropolitaines de recensement (RMR) telles que définies pour le recensement de 1981, produire des estimations mensuelles du chômage dont le coefficient de variation (CV) ne dépasserait pas 20%.
- (iii) Pour 66 régions économiques infraprovinciales choisies avec l'accord des provinces, produire des estimations mensuelles du chômage dont le CV ne dépasserait pas 25%.
- (iv) Pour les agglomérations de recensement de 60,000 habitants ou plus du Québec et de l'Ontario et de 25,000 habitants ou plus des autres provinces, produire des estimations trimestrielles moyennes du chômage dont le CV ne dépasserait pas 25%.

Departures from the overall strategy were made where necessary to accommodate specific user requirements. As an example of this, the province of Quebec expressed the need for monthly estimates for the CA of Sherbrooke, which according to the general strategy would have qualified only for three-month average estimates.

5.3 Steps in Allocating the Sample

To meet the data reliability objectives for improved subprovincial data under the redesigned sample, departure from the uniform nature of the old allocation was necessary. The allocation of the redesigned sample was carried out as described below.

5.3.1 Optimum relative sampling rates

As a preliminary step, the optimal relative sampling rates for NSR and SR areas under a proportional allocation scheme as it existed for the old design was determined using up-to-date data. In order to make the desired improvement in sub-provincial data, different sampling rates were needed from one sub-provincial area to the next. However, it was felt that, where possible, the optimal relative rates should serve as the basis for allocating the sample within sub-provincial areas.

A simple cost and variance model introduced by Fellegi, Gray and Platek (1967) during the 1962 redesign of the LFS was used. Under the model the total cost (C) and variance (V) are respectively given by

$$C = \frac{C_1 P_1}{W_1} + \frac{C_2 P_2}{W_2}$$

and

$$V = W_1 P_1 \sigma_1^2 + W_2 P_2 \sigma_2^2$$

where 1 and 2 denote area type (1 for SR, and 2 for NSR), C and σ^2 are unit (i.e., per person) cost and unit variance, P denotes population and $1/W$ is the sampling rate.

Fellegi et al. showed that to minimize costs while variances are fixed, the ratio of sampling rates is:

$$\frac{W_1'}{W_2'} = \frac{\sigma_2}{\sigma_1} \left(\frac{C_1}{C_2} \right)^{1/2}$$

Parameters for the model were estimated as follows:

- (i) Unit costs: Historical per dwelling costs by type of area were available. These were decreased by 10% for NSR areas to take into account the estimated effect of the shift from personal to telephone interviewing of households during their second through sixth months in the sample. The population was based on the 1981 Census.

On a fait des exceptions pour répondre à certaines exigences précises des utilisateurs. Ainsi, on a accédé à la demande de la province de Québec, dont les représentants ont dit avoir besoin d'estimations mensuelles relativement à l'agglomération de recensement de Sherbrooke, pour laquelle on ne devait produire en principe que des estimations trimestrielles moyennes.

5.3 Répartition de l'échantillon

Pour atteindre les objectifs de fiabilité des données et améliorer les données infraprovinciales au moyen du nouveau plan de sondage, il a fallu laisser tomber la répartition uniforme de l'échantillon qui avait caractérisé l'ancien plan. Le mode de répartition de l'échantillon prélevé selon le nouveau plan de sondage est décrit ci-après.

5.3.1 Fractions de sondage relatives optimales

L'étape préliminaire à la répartition a consisté à déterminer, au moyen de données à jour, les fractions de sondage relatives optimales pour les secteurs NAR et AR selon le mode de répartition proportionnelle employé dans l'ancien plan de sondage. Alors que pour apporter les améliorations voulues aux données infraprovinciales il aurait fallu appliquer des fractions de sondage différentes à chacune des régions infraprovinciales, on a estimé qu'on devrait, autant que possible, se servir des fractions de sondage relatives optimales pour répartir l'échantillon entre ces régions.

On s'est servi d'un modèle simple de coûts et de variances proposé par Fellegi, Gray et Platek (1967) pendant le remaniement de 1962. Selon ce modèle, le coût total (C) et la variance (V) sont obtenus respectivement par les formules

$$C = \frac{C_1 P_1}{W_1} + \frac{C_2 P_2}{W_2}$$

et

$$V = W_1 P_1 \sigma_1^2 + W_2 P_2 \sigma_2^2$$

où 1 et 2 indiquent le genre de secteur (1 pour AR et 2 pour NAR), C et σ^2 sont le coût par unité (c.-à-d. par personne) et la variance par unité, P est la population et $1/W$ est la fraction de sondage.

Fellegi et ses collaborateurs ont montré que, pour réduire les coûts au maximum lorsque les variances sont fixes, le ratio des fractions de sondage est

$$\frac{W_1'}{W_2'} = \frac{\sigma_2}{\sigma_1} \left(\frac{C_1}{C_2} \right)^{1/2}$$

On a estimé les paramètres du modèle de la manière suivante :

- (i) Coûts par unité : On connaissait les anciens coûts par logement selon le genre de secteur. On a diminué de 10% les coûts dans les secteurs NAR, pourcentage estimé de la différence engendrée par le fait que les interviews se feraient dorénavant par téléphone du deuxième au sixième mois ou les ménages seraient dans l'échantillon. Les chiffres de population étaient basés sur les résultats du recensement de 1981.

- (iii) Unit variances: Optimization was carried out for the characteristic unemployed, for which variances were given by:

$$\sigma_j^2 = F_j \frac{u_j}{P_j} (P_j - u_j),$$

where $j = 1, 2$ (area type), F = design effect for unemployed and u = persons unemployed.

Average design effects by type of area were determined from the old design and adjusted to reflect anticipated reductions due to the structural improvements in the redesigned sample. The factors for design effect reductions were obtained from studies carried out during the preliminary research phase of the sample redesign.

Persons unemployed, u , was derived by multiplying the average unemployment to population ratio for the period 1980-1982 by the 1981 Census population. The 1980-82 period was used, since average unemployment rates for the period conformed closely to medium term forecasts for the 1980s and early 1990s then being made by the Department of Finance and others. It included two years of comparatively low unemployment rates (7.5% in 1980 and 1981), and one year of higher unemployment (11% in 1982).

This analysis indicated that an allocation very close to proportional allocation of the sample to NSR and SR areas was optimal. It also indicated that the NSR areas, as suspected, were oversampled under the old design (see Columns (1)-(3), Table 1).

- (iii) Variances par unité : On a effectué l'optimisation en fonction de la caractéristique "chômeurs" pour laquelle les variances étaient définies :

$$\sigma_j^2 = F_j \frac{u_j}{P_j} (P_j - u_j),$$

où $j = 1, 2$ (genre de secteur), F = effet du plan pour les chômeurs et u = nombre de chômeurs.

On a déterminé les effets moyens du plan selon le genre de secteur d'après l'ancien plan de sondage et on les a corrigés à la baisse pour tenir compte des conséquences qu'auraient les améliorations structurelles apportées dans le nouveau plan. Des études effectuées pendant la phase de recherche qui a précédé le remaniement du plan de sondage ont permis de connaître les facteurs susceptibles de réduire les effets du plan.

On a obtenu u , le nombre de chômeurs, en multipliant le ratio moyen du nombre de chômeurs à la population pendant la période 1980-1982 par la population au recensement de 1981. On a pris la période 1980-1982 parce que les taux de chômage moyens pour cette période sont très proches des prévisions à moyen terme effectuées pour les années 80 et le début des années 90 par certains organismes, dont le ministère des Finances. On y a enregistré un taux de chômage relativement faible pendant deux ans (7.5% en 1980 et en 1981) et un taux élevé pendant un an (11% en 1982).

Selon les résultats de cette analyse, la répartition optimale était une répartition très proche de la répartition proportionnelle de l'échantillon entre les secteurs NAR et les secteurs AR, et les secteurs NAR étaient, comme on s'en doutait, suréchantillonnés dans l'ancien plan de sondage (voir les colonnes 1-3 du tableau 1).

Table 1
Percent of Sample in Self Representing Areas within Provinces

Province	Old Sample	Proportional Allocation	Optimum Allocation	Redesigned Sample
	Ancien plan	Répartition proportionnelle	Répartition optimale	Nouveau plan
Newfoundland - Terre-Neuve	41.8	51.3	42.6	44.6
Prince Edward Island - Île-du-Prince-Édouard	26.6	32.8	32.8	28.9
Nova Scotia - Nouvelle-Écosse	37.3	57.4	58.8	51.9
New Brunswick - Nouveau-Brunswick	49.5	52.5	47.4	53.6
Quebec	56.8	74.8	71.6	68.9
Ontario	62.5	79.1	78.8	75.0
Manitoba	54.1	71.0	76.4	56.4
Saskatchewan	44.7	51.8	62.1	56.8
Alberta	60.0	68.6	72.6	62.3
British Columbia - Colombie-Britannique	58.0	78.0	74.6	69.7
Canada	53.2	67.1	67.4	62.3

Tableau 1
Pourcentage de l'échantillon dans les secteurs AR, par province

5.3.2 Adjustments to the optimum allocation

To meet the reliability objectives for the smaller subprovincial areas as noted in section 5.2 and to realize the saving resulting from the structural improvements in the design, the following adjustments were made to the optimal allocation.

(i) Fix Sample Sizes for Smaller Subprovincial Regions

The sample sizes were adjusted for the smaller ERs and CMAs that would not have met targeted data reliability levels under an optimal-proportional allocation by type of area. To help this process, the data reliability targets were translated into standards for minimally acceptable sample sizes of 300 households per month for monthly estimates for CMAs and Economic Regions. Similarly, the standard adopted for three-month estimates for CAs was a sample size of 120 households or more per month. In contrast, under the old design, monthly data for smaller CMAs were published based on sample sizes of 130 households per month.

(ii) Allocation to Remaining Areas

The old provincial sample sizes, adjusted for smaller regions, as noted in (i) above, were allocated to various subprovincial regions in each province using the optimum relative sampling rates as determined in 5.3.1.

(iii) Sample Reduction due to Structural Improvements

Studies undertaken indicated that the structural improvements resulting from more efficient methods and procedures developed for the redesign would lead to a significant increase in the reliability of data at various levels of aggregation. The choice, therefore, was either to retain the increased data reliability or to translate these gains into savings. A decision was made to reduce the provincial sample sizes to the point where the reliability of provincial level estimates was the same as that under the old design. The decreases ranged from 5% to 9% depending on the province, and they were generally confined to the largest CMAs and ERs in an effort to achieve more uniformity in the reliability of subprovincial data.

This yielded a sample size for the redesigned LFS of 51,700 households, a 7% reduction over the sample size under the old design. It excludes a special sample of 1,300 dwellings in Alberta. This special sample is not part of the national base, and is funded by Alberta in order to obtain monthly estimates for medium sized cities in the province and annual estimates for Census Divisions.

5.3.2 Modifications apportées à la répartition optimale

Pour atteindre les objectifs de fiabilité dans les petites régions infraprovinciales dont il a été question à la section 5.2 et réaliser les économies qui devraient découler des améliorations structurelles apportées au plan, il a fallu modifier la répartition optimale de la manière suivante.

(i) Taille fixe des échantillons dans les petites régions infraprovinciales

On a modifié la taille des échantillons dans les petites RE et RMR pour lesquelles les données n'auraient pas atteint le niveau de qualité voulu si on avait fait une répartition optimale, proportionnelle selon le genre de région. Pour faciliter ce processus, on a établi des normes en matière de taille minimum de l'échantillon devant permettre d'atteindre l'objectif de fiabilité, en l'occurrence : 300 ménages par mois pour les estimations mensuelles relatives aux RMR et aux RE et 120 ménages ou plus par mois pour les estimations trimestrielles relatives aux agglomérations de recensement. Il convient de signaler à titre de comparaison que, selon l'ancien plan de sondage, les données mensuelles publiées sur les petites RMR étaient fondées sur un échantillon de 130 ménages par mois.

(ii) Répartition dans les régions restantes

On a réparti l'échantillon entre les régions infraprovinciales de chaque province en prenant les mêmes tailles d'échantillon que dans l'ancien plan, en apportant les modifications indiquées au paragraphe précédent pour les petites régions et en appliquant les fractions de sondage relatives optimales déterminées selon la méthode indiquée à la section 5.3.1.

(iii) Réduction de l'échantillon en raison des améliorations structurelles

Selon les résultats d'études menées à cet égard, les améliorations structurelles qui découleraient d'une plus grande efficacité des méthodes élaborées dans le cadre du remaniement du plan de sondage entraîneraient une augmentation marquée de la fiabilité des données à divers degrés d'aggrégation. Il s'agissait alors de choisir entre conserver cet avantage ou le remplacer par des économies d'argent. On a décidé de réduire la taille des échantillons provinciaux de manière à ce que le degré de fiabilité des estimations provinciales soit le même qu'avec l'ancien plan de sondage. Cette diminution a varié de 5 à 9% selon la province, et on l'a généralement effectuée dans les très grandes RMR et RE afin que la fiabilité des données infraprovinciales soit uniforme.

L'échantillon ainsi obtenu était de 51,700 ménages, soit une baisse de 7% par rapport à l'ancien plan de sondage. Ce chiffre ne tient pas compte d'un échantillon spécial de 1,300 ménages en Alberta. (Cet échantillon spécial n'est pas prélevé dans la base nationale, et c'est la province de l'Alberta qui le finance dans le but de disposer d'estimations mensuelles pour les villes de taille moyenne et d'estimations annuelles pour les divisions de recensement.)

(iv) **Post-Redesign Decrease in Sample Size**

Following the introduction of the redesigned sample in 1985, the sample size continued to be 51,700 until June 1986. At this time the sample size was reduced by 10% as a cost saving measure. The sample size decreases were concentrated in larger CMAs within each province, with 33% cuts in Montreal, Toronto, and Vancouver; 20%-25% in St. John's, Halifax, Saint John, Ottawa-Hull, Winnipeg, Regina, Saskatoon, Calgary, and Edmonton; and 10% in Quebec, Hamilton, and Victoria. Additionally, cuts of 20% occurred in certain Census Agglomeration areas.

This sample size reduction was achieved initially by means of sample size stabilization, whereby a portion of the selected dwellings was dropped prior to the interview. Later, changes were made in the basic stratum sampling rates to directly select the reduced size of the sample.

Sample sizes for the post-1971 Censal design (both before and after the 1976/77 sample size increase), and the post-1981 Censal design (initial, and post-1986 decrease) are given in Table 2.

(iv) **Diminution de la taille de l'échantillon subséquente au remaniement**

Après l'introduction de l'échantillon prélevé selon le nouveau plan en 1985, on en a maintenu la taille à 51,700 ménages jusqu'en juin 1986. A ce moment-là, on l'a réduit de 10% pour faire baisser les coûts. Ce sont les échantillons des grandes RMR dans chaque province qui ont subi une diminution : ceux de Montréal, de Toronto et de Vancouver, de 33%, ceux de St. John's, de Halifax, de Saint-Jean, d'Ottawa-Hull, de Winnipeg, de Regina, de Saskatoon, de Calgary et d'Edmonton, de 20 à 25% et ceux de Québec, de Hamilton et de Victoria, de 10%. L'échantillon de certaines agglomérations de recensement a également été réduit de 20%.

Au départ, on a diminué la taille de l'échantillon en stabilisant celle-ci, c'est-à-dire en laissant tomber une partie des logements choisis avant l'interview. Par la suite, on a modifié les fractions de sondage de base de la strate afin de déterminer directement la taille réduite de l'échantillon.

Le tableau 2 donne la taille des échantillons dans le plan remanié après le recensement de 1971 (avant et après l'élargissement de l'échantillon survenu en 1976-1977) et dans le plan remanié après le recensement de 1981 (initiale et subséquente à la réduction de 1986).

Table 2

**LFS Sample Sizes by Province
Old and New Designs**

Province	Old LFS Design		New LFS Design		
	Ancien plan de l'EPA		Nouveau plan de l'EPA		
	Pre-increase (1972-1976)	Post-increase (1976-1984)	Initial (1985-May 1986)	Post-decrease (June 1986)	Post-increase (April 1990)
	Avant l'élargissement (1972-1976)	Après l'élargissement (1976-1984)	Taille initiale (1985-Mai 1986)	Après la réduction (Jun 1986)	Après l'élargissement (Avril 1990)
Newfoundland - Terre-Neuve	1,815	3,091	2,789	2,628	2,970
Prince Edward Island - Île-du-Prince-Édouard	479	1,436	1,421	1,421	1,421
Nova Scotia - Nouvelle-Écosse	3,262	4,234	3,833	3,508	4,409
New Brunswick - Nouveau-Brunswick	2,539	4,260	3,854	3,497	3,843
Quebec	7,308	8,555	8,044	7,074	11,956
Ontario	9,544	10,903	10,246	9,176	18,045
Manitoba	1,860	4,489	4,139	3,676	4,297
Saskatchewan	1,997	5,989	5,453	4,927	4,963
Alberta	3,404	6,806	6,316	5,605	5,625
British Columbia - Colombie-Britannique	4,197	5,968	5,574	4,949	5,470
Canada	36,405	55,731	51,669	46,461	62,999

Tableau 2

**Taille de l'échantillon de l'EPA selon la province
Ancien et nouveau plans**

5.3.3 Impact of sample size reductions

As a result of improvements in the sample design in NSR and SR areas, and the improved allocation of the sample to the two types of area, it was possible to achieve substantial data reliability gains for sub-provincial data and comparable levels of data reliability at the province level, with a sample size 7% smaller than that prior to the redesign.

For the 13 of the 66 Economic Regions, 6 of the 24 CMAs, and 27 of the 42 non-CMA cities where sample size increases were required, an average 28% reduction in the CVs for the unemployed was obtained. In addition, for the 30 ERs with old CVs in the range of 15%-25%, there was an average 12% reduction in the CVs.

In addition to improvements in the quality of subprovincial data, the amount of subprovincial data was significantly increased under the redesigned sample. The number of ERs for which data are published increased from 53 to 67. Also, the number of CMAs for which data are published was increased from 23 to 24 as the result of the addition of one new CMA in the 1981 Census. Finally, the capability to produce three-month average estimates for the 42 CAs was added.

Table 3
Targeted Coefficients of Variation for Unemployed by Province, Old and New Designs

Province	Old Design	New Design	
	Ancien plan	Nouveau plan	
	Households (55,731)	Initial (51,669 households)	Post-decrease (46,461 households)
	Ménages (55,731)	Taille initiale (51,669 ménages)	Après la réduction (46,461 ménages)
Newfoundland – Terre-Neuve	7.2	7.3	7.5
Prince Edward Island – Ile-du-Prince-Edouard	9.7	8.9	8.9
Nova Scotia – Nouvelle-Ecosse	5.5	5.4	5.7
New Brunswick – Nouveau-Brunswick	6.3	6.4	6.6
Quebec	3.6	3.8	4.2
Ontario	3.7	3.7	4.0
Manitoba	6.5	6.8	7.5
Saskatchewan	7.2	7.3	7.6
Alberta	6.2	6.2	6.7
British Columbia – Colombie-Britannique	5.3	5.4	5.9
Canada	1.9	1.9	2.1

5.3.3 Conséquences de la réduction de la taille de l'échantillon

Grâce aux améliorations apportées au plan de sondage des secteurs AR et NAR et à la répartition de l'échantillon entre ces deux types de secteurs, il a été possible d'augmenter de façon marquée la fiabilité des données infraprovinciales et de maintenir à un niveau comparable la fiabilité des données provinciales tout en ayant un échantillon dont la taille était inférieure de 7% à celle de l'échantillon avant le remaniement.

Pour 13 des 66 régions économiques, 6 des 24 RMR et 27 des 42 villes autres que les RMR où il a fallu élargir l'échantillon, on a obtenu une diminution moyenne de 28% du CV des estimations relatives au chômage. En outre, on a réussi à faire baisser de 12% environ le CV pour les 30 RE où celui-ci se situait auparavant entre 15 et 25%.

Non seulement on a amélioré la qualité des données infraprovinciales produites au moyen du nouveau plan, mais on a aussi nettement augmenté leur quantité. Le nombre de RE pour lesquelles des données sont publiées est passé de 53 à 67, et le nombre de RMR, de 23 à 24, du fait qu'une nouvelle RMR a été ajoutée au recensement de 1981. Enfin, on s'est donné les moyens de produire des estimations trimestrielles moyennes pour 42 agglomérations de recensement.

Tableau 3
Coefficients de variation visés pour les estimations relatives au chômage, selon la province, Ancien et nouveau plans

Given the concentration of the post-redesign sample size decrease in the larger CMAs, the data reliability gains for the smaller sub-provincial regions were left intact. However, in addition to the reduced data reliability for the CMAs in question, the sample size decrease resulted in a 10% increase in the CV for the national estimate of unemployed, and similar reductions in data reliability for several of the provinces (see Table 3).

5.3.4 Sample size increase in 1989-90

During the period from November 1989 to April 1990, the LFS sample size was increased to provide improved estimates for Unemployment Insurance Commission regions. The increased sample was allocated only to regions which required it to yield estimates of unemployment with a coefficient of variation not exceeding 15% for quarterly estimates. Regions which already met this requirement were not changed. As a result, the bulk of the new sample fell in Ontario and Quebec. The last column of Table 2 presents the provincial sample sizes after the increase.

Comme la réduction de l'échantillon a touché principalement les grandes RMR, les gains réalisés sur le plan de la fiabilité des données pour les petites régions infraprovinciales sont demeurés intacts. Toutefois, cette réduction a entraîné non seulement une baisse de la fiabilité des données pour les RMR en question, mais aussi, comme on peut le voir au tableau 3, une augmentation de 10% du CV des estimations nationales relatives au chômage et une baisse semblable de la fiabilité des données pour plusieurs provinces.

5.3.4 Augmentation de la taille de l'échantillon en 1989-1990

Pour la période allant de novembre 1989 à avril 1990, la taille de l'échantillon de l'EPA a été augmentée afin d'améliorer les estimations pour les régions de la Commission d'assurance chômage. L'augmentation de la taille de l'échantillon a été allouée seulement aux régions qui le nécessitaient afin d'obtenir des estimations du chômage ayant un coefficient de variation n'excédant pas 15% pour les estimations trimestrielles. Ainsi, les régions qui respectaient déjà ce critère n'ont pas été modifiées. La majorité des nouvelles unités sont situées en Ontario et au Québec. La taille des échantillons pour les provinces (après l'augmentation) est présentée à la dernière colonne du tableau 2.

Chapter 6

Sample Rotation

6.1 Introduction

Rotation of the LFS sample involves complete or partial replacement of sampling units conducted in such a way that the sample for each month is a probability sample. Rotation of sampling units occurs at each stage of the multi-stage sample design. The ultimate unit of selection, the household, is replaced every six months, whereas successively higher staged units remain in the sample for longer periods of time. The determination of six months as the period for rotation of households is a trade off between the cost of rotation and the increase in non-response that might occur if respondents were asked to remain in the survey for a longer period of time.

To ensure uniform interviewer workloads and to minimize the effect of any bias due to the number of months a household has been in the survey, a rotation scheme was adopted whereby one sixth of the households rotate each month. This is achieved by associating with each PSU-Group-Cluster a rotation number between one and six. This number determines the months in which the rotation of households takes place (rotation number one implies the rotation of households in January and July; two in February and August; etc.). Described in the following sections are the methods by which rotation of units takes place, the operational steps involved in sample rotation, and the methods by which rotation numbers are assigned.

6.2 Method of Rotation

In the LFS design, sampling units at each stage of selection rotate. Although dwellings remain in the sample for only six months, units at successively higher stages of sampling remain in the sample for longer periods of time before rotating.

Rotation of the sample is relatively straightforward for the sampling methods followed in the design (randomized PPS systematic in the NSR area, random group method in the SRUs, and PPS systematic in apartment and special areas).

Chapitre 6

Renouvellement de l'échantillon

6.1 Introduction

Le renouvellement de l'échantillon de l'EPA consiste à remplacer une partie ou la totalité des unités d'échantillonnage, et il est effectué de manière qu'on ait chaque mois un échantillon probabiliste. On procède au renouvellement des unités à chaque étape du processus d'échantillonnage à plusieurs degrés. La dernière unité de sélection, le ménage, est remplacée tous les six mois, tandis que les unités sélectionnées aux degrés antérieurs demeurent dans l'échantillon pendant une période plus longue, les unités du premier degré y restant plus longtemps que celles du deuxième, et ainsi de suite. On a fixé à six mois la période de renouvellement de l'échantillon des ménages, car cette durée permettait d'atteindre un équilibre entre l'augmentation des coûts à laquelle on assisterait si le renouvellement était plus fréquent et l'accroissement du taux de non-réponse qui pourrait se produire si l'on demandait aux ménages de participer à l'enquête plus longtemps.

Pour assurer aux intervieweurs une charge de travail uniforme et réduire au maximum le biais associé au nombre de mois pendant lesquels un ménage a été interrogé, on a adopté un mode de renouvellement selon lequel un sixième de l'échantillon est remplacé chaque mois. Pour y arriver, on attribue à chaque UPE groupe/grappe un numéro de renouvellement entre un et six, qui détermine quel mois le ménage sera remplacé (tous les ménages portant le numéro de renouvellement un sont remplacés en janvier ou en juillet, ceux portant le numéro deux, en février ou en août, etc.). Les sections qui suivent décrivent la méthode de renouvellement des unités d'échantillonnage, le processus de renouvellement de l'échantillon et le mode d'attribution des numéros de renouvellement.

6.2 Méthode de renouvellement

Selon le plan de l'EPA, les unités sont renouvelées à chaque degré d'échantillonnage. Alors que les logements ne restent dans l'échantillon que pendant six mois, plus on remonte vers le premier degré d'échantillonnage, plus les unités demeurent dans l'échantillon longtemps.

Le renouvellement de l'échantillon est une opération relativement simple compte tenu des méthodes d'échantillonnage adoptées (méthode d'échantillonnage systématique avec classement aléatoire et ppt pour les secteurs NAR, méthode des groupes aléatoires pour les UAR et méthode d'échantillonnage systématique pour les immeubles d'appartements et les secteurs spéciaux).

In the two-stage SR design, rotation occurs as follows. Under the selection method, clusters within a stratum are randomly ordered and partitioned into groups. For the initially selected cluster within each random group, two numbers between one and the cluster inverse sampling ratio (ISR) are generated at random. The first determines a random start for systematic selection of dwellings within the cluster. The second determines the number of systematic samples of dwellings to be drawn from the cluster, that is, the number of six-month periods for which the cluster will remain in the sample.

Prior to each occasion for selecting a new sample of dwellings, the random start for the cluster is incremented by one, until the incremented value would exceed the cluster ISR, at which time the start reverts to 1. After the randomly determined number of sampling occasions, cluster rotation occurs.

The random retention period for initially selected clusters is necessary to ensure that initial probabilities of selection of units are preserved over time. If, for example, initially selected units were retained until exhausted (that is, until all systematic samples of dwellings were used), this would eventually result in a biased sample with overrepresentation of larger units. Cluster rotation is carried out by proceeding to the next cluster on the randomized list of clusters in the group. Replacements for initially selected units, and subsequent replacements, all remain in the sample until exhausted. As with initially selected clusters, the selection of dwellings is governed by a random start between 1 and the ISR which advances at each sampling occasion.

In NSR areas, rotation of dwellings and clusters proceeds similar to that in SR areas, and additionally there is rotation of rural groups. The urban groups of the PSUs, since they are selected with certainty, do not rotate but remain in the sample as long as the PSU does. Rotation of rural groups and PSUs is similar to rotation of clusters, except that, for both units, minimum periods of time are established before rotation takes place, due to the costs associated with their rotation. For rural groups this period corresponds to the minimum isr of all the rural groups within the PSU, whereas for PSUs the minimum period is staggered across each province; this guards against any discontinuity in the time series due to rotating many units at once. Because the rotation of initially selected units is delayed in this fashion, the rotation of replacement units is correspondingly advanced to preserve the correct probabilities of selection over time.

Le renouvellement se fait de la manière suivante pour les secteurs AR, où l'échantillonnage est à deux degrés. Selon la méthode de sélection, les grappes d'une strate sont classées au hasard et réparties en groupes. Pour la première grappe choisie dans chaque groupe aléatoire, on génère au hasard deux nombres entre 1 et la fraction de sondage inverse (fsi) de la grappe. Le premier détermine l'origine aléatoire pour l'échantillonnage systématique des logements dans la grappe. Le second détermine le nombre d'échantillons systématiques de logements à prélever dans la grappe, c'est-à-dire le nombre de périodes de six mois (cycles) pendant lesquelles la grappe fera partie de l'échantillon.

Avant chaque prélèvement d'un nouvel échantillon de logements, on ajoute 1 à l'origine aléatoire de la grappe, jusqu'à ce que la valeur obtenue dépasse la fsi de la grappe; à ce moment-là, l'origine redevient 1. Une fois qu'on a atteint le nombre de prélèvements de l'échantillon qui a été choisi au hasard, on procède au remplacement de la grappe.

Il est nécessaire d'attribuer à chaque première grappe choisie un nombre aléatoire de cycles pendant lesquels elle restera dans l'échantillon pour maintenir la probabilité initiale de sélection des unités. Si, par exemple, on gardait les premières unités choisies jusqu'à épuisement de l'échantillon, c'est-à-dire jusqu'à ce que tous les échantillons systématiques de logements aient été prélevés, l'échantillon finirait par devenir biaisé, car les grandes unités y seraient sur-représentées. On procède au renouvellement d'une grappe en passant à la grappe suivante dans la liste randomisée de grappes par groupe. Les unités qui remplacent les premières unités choisies et toutes les unités remplaçantes qui suivent restent dans l'échantillon jusqu'à ce qu'on ait épuisé celui-ci. Comme dans le cas des premières grappes choisies, le choix des logements est déterminé par une origine aléatoire entre un et la fsi, à laquelle on ajoute 1 à chaque prélèvement de l'échantillon.

Pour les secteurs NAR, on procède au renouvellement des logements et des grappes comme pour les secteurs AR, mais il y a en plus le renouvellement des groupes ruraux. Comme les groupes urbains des UPE sont choisis avec une probabilité égale à 1, ils ne sont pas remplacés et restent dans l'échantillon aussi longtemps que l'UPE. Le renouvellement des groupes ruraux et des UPE est semblable à celui des grappes, sauf que, dans les deux cas, la période minimum de conservation dans l'échantillon est établie avant que le renouvellement ait lieu, à cause des coûts associés à ce dernier. Pour les groupes ruraux, cette période correspond à la fsi minimum de tous les groupes ruraux dans l'UPE, tandis que, pour les UPE, elle est échelonnée à l'intérieur de chaque province, pour éviter la discontinuité dans les séries chronologiques qui résulterait du renouvellement d'un trop grand nombre d'unités à la fois. Le renouvellement des premières unités étant ainsi retardé, le renouvellement des unités de remplacement est avancé en conséquence pour maintenir les bonnes probabilités de sélection.

6.3 Operational Steps Involved in Sample Rotation

Rotation of the sample is automated within a system known as the Sample Design System, which identifies the units that are rotating in and out of the sample for each survey. These rotating units require manual processing by Sample Control staff, who identify the geographical area represented, and form lower stage sampling units where applicable.

Each of the cluster selection computer programs create *rotation records*, which describe the pattern of rotation after each initial cluster selection. These rotation records describe the sequence of advancing starts within clusters and the sequence of between cluster rotations corresponding to each selection within a group. Each record lasts the life of the group, up to a maximum of 40 random starts. The collection of these records constitutes the master rotation file, which is used to automatically rotate the sample. The master rotation file is changed as rotation patterns are updated (as in the open-ended apartment frame), and as old records become exhausted and are replaced by new ones.

The way in which the rotation of households is accomplished is as follows. Seven months before a particular survey date, a file is created that contains design information on all clusters newly rotating into the sample for that survey date. This file is used to initiate the database for listing purposes, and to identify for Sample Control the clusters for which cluster diagrams (F01s) are required. The F01s are sent to the Regional Offices (ROs) to be received coincident with the Sample Status Report, which is generated 22 to 23 weeks before the date of introduction and identifies all the new clusters requiring listing. The process of listing requires that a separate list line be completed on the Cluster List (F02) for each dwelling in the cluster.

The lists are entered on minicomputers in the ROs and are transmitted to the central database. Prior to the survey date a file (which is created from the rotation files by the Sample Design System) is fed into the database; this file specifies the random start and inverse sampling ratio to be applied to each cluster in the sample. Thereafter, selection of households is accomplished by a systematic sample of the lists, and household dockets are printed in the RO for each selected household. Households remain in the sample for six months (or less if they are introduced in a month other than that associated with their rotation number, as when phasing in new parts of the sample). After the six months another set of random starts is reserved on the database, thereby resulting in selection of new households.

6.3 Processus de renouvellement de l'échantillon

Le renouvellement de l'échantillon se fait automatiquement au moyen du "Système de conception de l'échantillon", qui détermine quelles unités doivent être introduites dans l'échantillon de chaque enquête et lesquelles doivent en être supprimées. Ces unités sont traitées manuellement par le personnel de la Section du contrôle et de l'établissement des échantillons, qui doit déterminer quelles régions géographiques sont représentées et former, s'il y a lieu, les unités des degrés suivants d'échantillonnage.

Chaque programme informatique de sélection des grappes crée des *enregistrements de renouvellement* qui décrivent le plan de renouvellement suivi après le prélèvement de chaque grappe initiale. Ces enregistrements de renouvellement indiquent l'ordre dans lequel on ajoute 1 aux origines aléatoires pour l'échantillonnage des logements dans les grappes et l'ordre dans lequel les grappes prélevées dans un groupe se succèdent. Chaque enregistrement est valide tant que le groupe est soumis à l'échantillonnage, jusqu'à concurrence de 40 origines aléatoires. Les enregistrements réunis constituent le fichier principal de renouvellement, qui sert au renouvellement automatique de l'échantillon. Le fichier est modifié chaque fois que les plans de renouvellement sont mis à jour (comme pour la base ouverte d'appartements) et au fur et à mesure que les enregistrements deviennent périmés et doivent être remplacés.

Le renouvellement des ménages se fait de la manière suivante. Sept mois avant la date de l'enquête, on crée un fichier dans lequel on met tous les renseignements relatifs au plan de sondage des grappes qui seront introduites dans l'échantillon pour cette date. Ce fichier sert à la création d'une base de données aux fins de l'établissement des listes et indique au personnel de la Section du contrôle et de l'établissement des échantillons les grappes pour lesquelles il faut produire des schémas (F01). Les F01 sont envoyées aux bureaux régionaux (BR) et doivent y arriver en même temps que le Rapport sur l'état de l'échantillon, qui est lui-même produit 22 ou 23 semaines avant la date d'introduction des unités dans l'échantillon et qui indique toutes les nouvelles grappes pour lesquelles il faut établir une liste. Cette opération consiste à inscrire sur une ligne distincte de la Liste de la grappe (F02) les renseignements relatifs à chaque logement faisant partie de la grappe.

Les listes sont introduites dans les mini-ordinateurs des BR et sont transmises à la base de données centrale. Avant la date de l'enquête, on introduit dans la base de données un fichier, créé à partir des enregistrements de renouvellement par le Système de conception de l'échantillon, qui contient l'origine aléatoire et la fraction de sondage inverse à attribuer à chaque grappe de l'échantillon. La sélection des ménages est ensuite effectuée par échantillonnage systématique des listes, et un dossier est imprimé pour chaque ménage choisi. Les ménages font partie de l'échantillon pendant six mois (ou moins s'ils y ont été introduits au cours d'un mois autre que celui qui correspond à leur numéro de renouvellement, par exemple dans les régions où un nouvel échantillon est introduit progressivement), à la suite de quoi une nouvelle série d'origines aléatoires est réservée dans la base de données et de nouveaux ménages sont échantillonnés.

In NSR areas, a rotation record lasts only as long as a group remains in the sample. Thus, as records become exhausted, it can be determined which groups are exhausted. Then, by referring to the group frame, replacement groups are determined. In the groups identified as rotating in, Sample Control determines their geographical location, and prepares maps of the area. Ordinarily, the process of group rotation begins at least 30 weeks before the date of introduction of the sample.

With the rotation of PSUs, the subsequent stages of selection, namely groups and clusters, will have to be carried out, as outlined in Chapter 4. If the area was previously clustered, the same random ordering of units is used, although a different systematic sample is selected. Approximately the same length of time is required for replacement of PSUs, since the formation and selection of groups is not very time consuming.

6.4 Assigning Rotation Numbers

In assigning rotation numbers, the objective is to evenly distribute the sample in terms of expected take. This distribution is achieved simultaneously for the sample as a whole, and at the same time for the smallest possible geographical subsets of it. Adherence to these objectives implies the following: (i) the workload of interviewers is stable, as roughly equal numbers of units are rotating each month; (ii) the sample is comprised of equal numbers of households having been in the sample for 1 vs 2... vs 6 occasions, nullifying time-in-sample effects as a cause of differences in estimates between areas, or over time; and (iii) the sample is effectively divided into six equally representative parts, which may be used when sub-samples from the LFS frame are desired.

To achieve the aforementioned objectives in NSR areas, selected clusters were assigned rotation numbers in such a manner as to balance the total expected take within each PSU, stratum and Economic Region. The assignment of rotation numbers was accomplished independently within each Economic Region, but a factor was built in to randomly distribute that expected take that remained after equal amounts were assigned to each rotation number. Due to the randomness of the procedure, the unbalanced expected takes at the regional level should cancel out at higher levels.

Clusters in urban PSUs or urban portions of PSUs are blocks or block faces, with typically more than six selected per PSU, and with an average sample yield of 3 dwellings. Clusters in rural PSUs or rural portions of PSUs correspond to Census Enumeration Areas, which in some cases have undergone splitting or combining. For rural PSUs, frequently six clusters are formed with all six selected, whereas for rural portions of mixed PSUs typically two to three clusters, each with sample yields of 10 to 15 dwellings, are selected.

Dans les secteurs NAR, l'enregistrement de renouvellement demeure valide aussi longtemps que le groupe reste dans l'échantillon. On peut ainsi déterminer quels groupes sont épuisés d'après les enregistrements périmés et connaître les groupes de remplacement en consultant la base de groupes. La Section du contrôle et de l'établissement des échantillons détermine l'emplacement géographique des groupes devant être introduits dans l'échantillon et trace la carte des secteurs où ils se trouvent. En général, le renouvellement des groupes débute au moins 30 semaines avant la date d'introduction de l'échantillon.

Après le renouvellement des UPE, il faut procéder à l'échantillonnage des unités suivantes, soit les groupes et les grappes, selon la méthode décrite au chapitre 4. S'il s'agit d'un secteur qui a déjà été divisé en grappes, on utilise le même classement aléatoire des unités mais on prélève un échantillon systématique différent. Il faut prévoir autant de temps que pour le remplacement des UPE, car la formation et la sélection des groupes ne sont pas des opérations très longues.

6.4 Attribution des numéros de renouvellement

Le but poursuivi lorsqu'on attribue les numéros de renouvellement est de répartir l'échantillon de telle manière que le rendement prévu soit le même partout, tant au niveau de l'échantillon global que des plus petites unités géographiques. Si l'on respecte cet objectif, il en découle trois conséquences : (i) la charge de travail des intervieweurs est stable, étant donné qu'un nombre à peu près égal d'unités est remplacé chaque mois; (ii) l'échantillon est constitué d'un nombre égal de ménages qui en font partie depuis 1, 2, 3, 4, 5 ou 6 mois, ce qui annule l'effet que le nombre de mois passés dans l'échantillon pourrait avoir sur les estimations relatives aux différents secteurs et à différentes périodes; (iii) l'échantillon est bien divisé en six parties également représentatives, dont on peut se servir lorsqu'on veut obtenir des sous-échantillons de l'EPA.

Pour atteindre cet objectif dans les UNAR, on a attribué des numéros de renouvellement aux grappes échantillonnées de manière à ce que le rendement prévu de l'échantillon soit le même dans chaque UPE, strate et région économique. L'attribution des numéros de renouvellement s'est faite de façon indépendante dans chaque région économique, mais on a introduit un facteur permettant de répartir aléatoirement le rendement prévu qui restait une fois qu'on avait attribué un nombre égal d'unités aux grappes portant chaque numéro de renouvellement. En raison du caractère aléatoire de cette opération, les différences par rapport au rendement prévu au niveau de la région devraient s'annuler aux niveaux supérieurs.

Dans les UPE urbaines ou la partie urbaine des UPE, les grappes sont des îlots ou des côtés d'îlot. On en prélève généralement plus de six par UPE, dont le rendement prévu de l'échantillon est de 3 logements en moyenne. Dans les UPE rurales ou les parties rurales des UPE, les grappes correspondent à des secteurs de dénombrement du recensement, dont certains ont été divisés ou combinés à d'autres. Dans les UPE rurales, on forme souvent six grappes qui sont toutes échantillonnées, tandis que dans les parties rurales des UPE mixtes, on prélève généralement 2 ou 3 grappes ayant un rendement de l'échantillon de 10 à 15 logements.

Due to the differences between the size and numbers of sampled units in urban and rural areas, the first step in balancing the sample take by rotation was to balance it as much as possible for the larger rural units at the stratum or Economic Region level. Rotation numbers were then assigned to the urban units to balance the combined rural and urban take at the PSU level, as well as the rural and urban separately and combined at the stratum and Economic Region levels.

Initial assignment of rotation numbers for the NSR areas took into consideration the entire sample within each Economic Region. A running count of the sample take by urban and rural parts of PSUs, strata and Economic Regions is maintained, and when PSU rotation occurs, rotation numbers are assigned first to rural and then to urban units, so as to maintain the balance in the sample take across rotations simultaneously to the degree possible at the PSU, stratum and ER levels.

In SR areas every sub-unit has a multiple of six selections so that each rotation group can have the same number of selections. The total expected take is balanced by areas consisting of individual large cities (e.g., Toronto, Montreal) or groups of smaller cities within a province. Again, a random procedure to start the balancing of each area ensures that the provincial totals would also be balanced.

In the Apartment frame, every set of six selections in the systematic sample was assigned the six rotation numbers in a random order. No balancing is possible in this open-ended frame. For new additions to the list, a set of six randomly ordered rotation numbers is generated as required. The assignment of rotation numbers would entail choosing the next rotation from this set.

In the Special Area frame, the number of selections is so small that only a provincial balancing of the entire frame was attempted. In this frame the occasional group selection may be clustered but all cluster selections within the group have the same rotation number.

Comme la taille des unités échantillonnées et leur nombre ne sont pas les mêmes dans les régions rurales et dans les régions urbaines, la première chose à faire pour uniformiser le rendement de l'échantillon à chaque renouvellement a été d'essayer de faire en sorte que celui-ci soit aussi uniforme que possible dans les grandes unités rurales au niveau de la strate ou de la RE. On a ensuite attribué les numéros de renouvellement aux unités urbaines en essayant d'uniformiser le rendement de l'échantillon des régions rurales et urbaines combinées au niveau des UPE ainsi que des régions rurales et urbaines séparées et combinées au niveau de la strate et de la RE.

Au moment d'attribuer les premiers numéros de renouvellement aux secteurs NAR, on a tenu compte de l'échantillon entier dans chaque RE. On calcule le total cumulatif du rendement de l'échantillon selon les parties urbaines et les parties rurales des UPE, des strates et des RE et, lorsque le renouvellement des UPE a lieu, on attribue des numéros de renouvellement aux unités rurales d'abord et aux unités urbaines ensuite, de manière à garder, autant que possible, le rendement de l'échantillon égal en même temps au niveau de l'UPE, de la strate et de la RE.

Dans les secteurs AR, on prélève six ou un multiple de six sous-unités de manière à ce qu'il y ait le même nombre de sous-unités dans chaque groupe de renouvellement. Le rendement prévu total est uniformisé au niveau de régions constituées soit d'une grande ville (p.ex. Toronto ou Montréal), soit d'un groupe de petites villes d'une même province. On adopte encore une méthode à caractère aléatoire pour amorcer l'uniformisation du rendement de l'échantillon dans chaque région de manière à ce que le rendement total dans les provinces soit le même.

Dans la base d'appartements, on a attribué au hasard les six numéros de renouvellement à chaque série de six unités choisies. Il n'est pas possible d'uniformiser le rendement de l'échantillon dans cette base ouverte. Pour les nouvelles unités qui y sont ajoutées, on génère au besoin une série de six numéros de renouvellement aléatoires. Pour attribuer les numéros de renouvellement, il suffit de passer au numéro suivant de la série.

Dans la base de secteurs spéciaux, le nombre d'unités échantillonnées est si petit qu'on essaie seulement d'uniformiser le rendement de l'échantillon de la base entière au niveau provincial. Il arrive à l'occasion qu'un groupe choisi soit divisé en grappes, mais alors on attribue le même numéro de renouvellement à toutes les grappes qui y sont échantillonnées.

Chapter 7

Survey Operations

7.1 Introduction

This chapter outlines the activities and organization of survey operations. The first section identifies the organizational areas involved in survey operations and the second gives a schedule of main activities for a typical survey. The third and fourth sections deal with sample operations and data collection activities. The last section describes data processing and dissemination.

7.2 Organization and Responsibility

This description of survey organization includes those areas directly involved with the monthly operation of the LFS process, as well as areas that influence the survey in important ways, such as changing the survey to meet user requirements and Statistics Canada policy. Except for the Regional Offices (ROs), all areas are located in Head Office in Ottawa.

The Household Surveys Division within Labour and Household Surveys Branch is responsible for the overall operational control of the LFS. This includes setting the questionnaire, processing (edit, imputation, coding) the data, tabulating and analysing the monthly results and providing client consultation services. The Labour and Household Surveys Analysis Division, in the same Branch, deals with the detailed analysis and continual review of labour market developments as revealed by the survey data. It also determines and maintains survey concepts and content and engages in further data development by sponsoring supplementary surveys and analysis.

The Social Survey Methods Division has overall responsibility for the methodology of all household surveys. In particular, the division is responsible for directing and conducting the research and development of the LFS methodology. This involves carrying out all sample design work (e.g., sampling, rotation, updating, estimation, variance estimation and related methodological developments, development of evaluation programs and controls at different stages of the operation of the survey, maintenance of the quality of the design over time and preparation of sampling material for field operations).

Chapitre 7

Opérations de l'enquête

7.1 Introduction

Ce chapitre décrit les activités menées dans le cadre des opérations de l'enquête et l'organisation de ces opérations. La première section indique quelles divisions participent à ces opérations et la deuxième donne l'ordre dans lequel les principales activités sont accomplies pour une enquête typique. Les deux sections suivantes portent sur l'échantillonnage et la collecte des données, et la dernière, sur le traitement et la diffusion des données.

7.2 Organisation et responsabilités

Dans cette description de l'organisation de l'enquête, nous parlerons des unités administratives qui participent directement aux activités mensuelles de l'EPA ainsi que de celles qui ont une influence importante sur l'enquête, soit en déterminant les changements à y apporter pour mieux répondre aux besoins des utilisateurs, soit en définissant l'orientation du Bureau. À l'exception des bureaux régionaux, toutes ces unités administratives sont situées au bureau central à Ottawa.

La Division des enquêtes-ménages, qui fait partie de la Direction des enquêtes sur le travail et auprès des ménages, est chargée de voir à ce que l'ensemble des opérations de l'enquête soient menées à bien. Cela comprend l'élaboration du questionnaire, le traitement des données (contrôle, imputation, codage), la totalisation des données, l'analyse des résultats mensuels et les services de consultation auprès des clients. La Division de l'analyse des enquêtes sur le travail et les ménages, qui fait partie de la même direction, s'occupe de l'analyse détaillée et de l'examen continu de l'évolution du marché du travail révélée par les données de l'enquête, de la définition des concepts et de la détermination du contenu ainsi que de leur mise à jour s'il y a lieu et enfin de la production d'autres données par la tenue d'enquêtes supplémentaires et l'analyse de leurs résultats.

La Division des méthodes d'enquêtes sociales a la responsabilité de la méthodologie de toutes les enquêtes-ménages. En ce qui concerne l'EPA, elle est chargée de diriger et de réaliser la recherche et le développement ayant trait à la méthodologie de cette enquête. Ces tâches comprennent tout le travail lié à l'élaboration du plan de sondage, c'est-à-dire à l'échantillonnage, au renouvellement et à la mise à jour de l'échantillon, à la production des estimations et à l'estimation de la variance, ainsi que les travaux méthodologiques connexes, l'élaboration des programmes d'évaluation et des contrôles appliqués à diverses étapes de la réalisation de l'enquête, le maintien de la qualité du plan de sondage et la préparation des documents nécessaires à l'échantillonnage pour les opérations sur le terrain.

Field work related to data collection is the responsibility of the Survey Operations Division in the Regional Operations Branch. The Head Office is responsible for the functional management of the data collection and capture activities. This involves supplying the Regional Offices with all necessary survey forms; generating training manuals, exercises, survey guides and information packages; conducting training sessions; and scheduling the data collection, capture and transmission operations.

Actual data collection is carried out by a team of 800 interviewers working under the direct supervision of eight Regional Offices located in St. John's, Newfoundland; Halifax, Nova Scotia; Montreal, Quebec; Sturgeon Falls, Ontario; Toronto, Ontario; Winnipeg, Manitoba; Edmonton, Alberta and Vancouver, British Columbia. These offices are responsible for the field work related to the LFS, as well as the Census, and many other household, institutional and business surveys. Their duties include recruitment, hiring and training of interviewers, listing of selected clusters and related map work, data collection, data capture and maintenance of quality control over these functions.

The Informatics Services and Development Division in the Informatics Branch oversees the network of minicomputers located in each RO and Head Office. This system transmits data for pre-printing survey forms and for receiving survey data. The division is also responsible for maintaining the majority of the software and mainframe computer operations related to all aspects of the survey, from sampling to weighting and tabulation.

Because no one area is responsible for all aspects of the survey, several coordinating groups are needed to (1) ensure that issues are brought to the attention of the appropriate area and (2) help the exchange of information between these areas. This coordination is provided through the following committees, which meet regularly.

1. Operations Committee – chaired by a member of the Household Surveys Division.
2. Content Consultation Group – chaired by a member of the Labour and Household Surveys Analysis Division.
3. Data Quality Committee – chaired by a member of the Social Surveys Methods Division.
4. Redesign Committee – chaired by a member of the Social Survey Methods Division.

Les opérations sur le terrain ayant trait à la collecte des données relèvent de la Division des opérations des enquêtes, qui fait partie de la Direction des opérations régionales. Le bureau central assume la direction fonctionnelle de la collecte et de la saisie des données. Il doit notamment fournir aux bureaux régionaux toutes les formules d'enquête nécessaires, produire les manuels de formation, les exercices, les guides et les troupes d'information, offrir la formation et établir le calendrier des opérations de collecte, de saisie et de transmission des données.

La collecte-même des données est effectuée par une équipe de 700 intervieweurs qui sont directement supervisés par les huit bureaux régionaux situés respectivement à St. John's (Terre-Neuve), Halifax (Nouvelle-Écosse), Montréal (Québec), Sturgeon Falls (Ontario), Toronto (Ontario), Winnipeg (Manitoba), Edmonton (Alberta) et Vancouver (Colombie-Britannique). Ces bureaux sont chargés de l'accomplissement du travail sur le terrain pour l'EPA, mais aussi pour le recensement et un grand nombre d'autres enquêtes menées auprès des ménages, des établissements institutionnels et des entreprises. Parmi leur fonctions, signalons celles de recruter, d'embaucher et de former les intervieweurs, d'établir les listes pour les grappes échantillonnées et de préparer les cartes s'y rapportant, de recueillir et de saisir les données et enfin de contrôler la qualité de ces activités.

La Division des services et du développement informatiques, à la Direction de l'informatique, s'occupe du réseau de mini-ordinateurs, lesquels sont situés dans chaque BR et au bureau central et ont pour fonction de transmettre les données, d'imprimer certains renseignements sur les formules d'enquête et de recevoir les données d'enquête. La Division est aussi chargée d'effectuer la plupart des activités informatiques, du point de vue des logiciels comme de l'ordinateur central, qui ont trait à tous les aspects de l'enquête, de l'échantillonnage à la pondération et à la totalisation des données.

Comme il n'y a pas une unité administrative qui a la responsabilité de tous les aspects de l'enquête, on a besoin de plusieurs groupes de coordination pour veiller à ce que les dossiers soient acheminés aux unités compétentes et pour faciliter l'échange d'information entre celles-ci. Les comités suivants, qui se réunissent à intervalles réguliers, remplissent cette fonction.

1. Comité des opérations – présidé par un membre de la Division des enquêtes-ménages.
2. Groupe de consultation sur le contenu – présidé par un membre de la Division de l'analyse des enquêtes sur le travail et les ménages.
3. Comité de la qualité des données – présidé par un membre de la Division des méthodes d'enquêtes sociales.
4. Comité du remaniement – présidé par un membre de la Division des méthodes d'enquêtes sociales.

7.3 Survey Schedule

The sequence of operations can be understood by examining the activities as they relate to the survey process for a particular month.

First, all clusters that are selected or are soon to be selected, have lists of dwellings which are maintained on the main computer in Ottawa. Five to six weeks prior to survey week the sample is selected. This entails identifying the dwellings associated with random starts and systematic samples in all clusters that are active for that particular survey. Specification of the sample dwellings is sent to the ROs to prepare the interviewer assignments.

Three weeks prior to survey week, information for all selected dwellings is sent to the ROs. For dwellings entering the survey for the first time (i.e., births), this consists only of control information. For the five-sixths of the sample dwellings that were in the survey during the previous month, this consists of control information as well as selected information from the survey forms from the previous month. The following week (two weeks prior to survey week) packages of survey materials are prepared by RO staff and shipped to the interviewers. The week prior to survey week is used by interviewers to plan their work. This includes scheduling when interviews will be conducted during the next week, and planning a route to follow for conducting face-to-face interviews. Interviews are conducted from Monday to Saturday of survey week and interviewers are instructed to return all completed survey documents to ROs daily.

Regional Office processing of returned survey documents begins on the Tuesday of survey week. Forms are batched, sent for data capture and capture verification on the minicomputer network. The completed batches are then transmitted to Ottawa. Transmissions occur nightly, with the final cut-off usually set as the Wednesday night of the week after survey week.

Editing and coding of the survey data begins on the Wednesday of survey week. This processing, done at Head Office, continues until the Thursday of the week after survey week, with weighting and tabulation done on the Friday. The steps involved in Head Office processing are further described in section 7.7.

Text for publication and dissemination is prepared and translated between the Tuesday and Thursday of the second week following survey week. Data for the particular survey is first released at 7 a.m. on the Friday of this week.

7.3 Calendrier de l'enquête

On peut voir dans quel ordre les opérations se déroulent en examinant les activités liées à la tenue de l'enquête qui sont accomplies au cours d'un mois.

Tout d'abord, pour chaque grappe échantillonnée ou sur le point de l'être, il y a une liste de logements qui est tenue à jour dans l'ordinateur principal à Ottawa. Entre cinq et six semaines avant la semaine d'enquête, on prélève l'échantillon. Il s'agit pour cela de déterminer quel logement correspond à l'origine aléatoire puis de former un échantillon systématique dans chaque grappe qui est active ce mois-là. Tous les détails concernant les logements de l'échantillon sont envoyés aux BR afin qu'on y prépare les tâches des intervieweurs.

Les renseignements sur les logements échantillonnés sont expédiés aux BR trois semaines avant la semaine d'enquête. Pour les logements introduits dans l'échantillon ce mois-là (les nouveaux dossiers), il s'agit de données de contrôle. Pour les autres (les cinq sixièmes des logements échantillonnés), il s'agit de données de contrôle et de renseignements tirés des formules d'enquête du mois précédent. La semaine suivante, soit deux semaines avant la semaine d'enquête, les documents d'enquête sont réunis par le personnel des BR et expédiés aux intervieweurs. Les intervieweurs consacrent la semaine précédant la semaine d'enquête à la planification de leur travail. Notamment, ils fixent l'horaire des interviews et tracent l'itinéraire qu'ils suivront pour effectuer celles-ci. Les interviews sont menées du lundi au samedi de la semaine d'enquête et les intervieweurs doivent retourner les documents remplis au BR au jour le jour.

Le dépouillement des documents d'enquête reçus au BR commence le mardi de la semaine d'enquête. Les formules sont mises en lots et envoyées pour que les données soient saisies et la saisie, vérifiée, sur mini-ordinateur, après quoi les lots sont transmis à Ottawa. La transmission se fait tous les soirs jusqu'à la date limite, habituellement fixée au mercredi soir de la semaine suivant la semaine d'enquête.

La vérification et le codage des données d'enquête débutent le mercredi de la semaine d'enquête. Ces opérations, qui se déroulent au bureau central, se poursuivent jusqu'au jeudi de la semaine suivant la semaine d'enquête, et sont suivies de la pondération et de la totalisation des données qui ont lieu le vendredi. La section 3.5 contient plus de détails sur les étapes du traitement des données qui se déroulent au bureau central.

Le texte qui sera publié et diffusé est rédigé et traduit entre le mardi et le jeudi de la deuxième semaine suivant la semaine d'enquête. La première diffusion des données tirées de l'enquête a lieu à 7 heures le vendredi de cette même semaine.

7.4 Sampling Activities

The Sample Control unit of Social Survey Methods Division maintains the survey sample. This involves (1) identifying in advance those sample units that will be rotating into and out of the sample; (2) forming the necessary sample units (i.e., clusters); (3) preparing the necessary maps for use by interviewers; and (4) identifying the systematic samples of dwellings from within clusters each month.

Counting

Approximately a year in advance, the Sample Control unit identifies new NSR Primary Sampling Units to be introduced. Maps of the PSUs, which show their locations, are sent to the Regional Offices to allow the ROs to determine if and where new interviewers might have to be hired.

The most recent Census data on counts and plotting of dwellings are used to sub-divide the PSUs into clusters of required size. If counts or plots are not adequate, additional maps of the area are prepared and sent to the Regional Offices for field counting. Field counting consists of visiting the area and identifying all dwellings within the boundaries. These are shown on the maps as counts of dwellings between obvious permanent geographic features (rivers, roads etc.). Once the count maps are received by the Sample Control Unit, the areas are divided into clusters.

Clustering

Clustering consists of dividing a large area into smaller areas according to criteria related to optimum density factors and sampling rates within the stratum. In urban places, clusters consist of blocks, groups of blocks or block faces. In rural areas clusters are usually Census Enumeration Areas (EAs). Where EA dwelling counts are too large or too small, EAs are split or combined to ensure enough clusters to represent all six rotation groups.

About six months before interviewing, all clusters that are to be introduced, in both SR and NSR areas, are identified. Clusters are sketched on Cluster Diagrams (Form 01), which have control information pre-printed on them. The Cluster Diagrams are then shipped to the ROs, where they are matched with pre-printed Listing Forms (Form 02) and sent to interviewers for listing the dwellings.

Listing

During the listing procedure each dwelling is assigned a unique listing line number and is recorded on the Form 02 by its street address or a description. All dwellings that can be determined as being habitable (whether vacant or occupied) and that are within the boundaries as outlined on the Cluster Diagram, are

7.4 Activités d'échantillonnage

La Sous-section du contrôle de l'échantillon de la Division des méthodes d'enquêtes sociales s'occupe de la mise à jour de l'échantillon. C'est là qu'on détermine d'avance quelles unités seront introduites dans l'échantillon et lesquelles en seront supprimées, qu'on forme les unités d'échantillonnage nécessaires (les grappes), qu'on prépare les cartes dont les intervieweurs auront besoin et qu'on constitue chaque mois les échantillons systématiques de logements dans les grappes.

Dénombrement

On détermine environ un an à l'avance les nouvelles UPE NAR à introduire dans l'échantillon. On envoie des cartes des UPE montrant leur emplacement aux BR pour permettre à ces derniers de décider s'il faudra embaucher de nouveaux intervieweurs et, le cas échéant, dans quelles régions.

On se sert des données du dernier recensement sur le nombre de logements et leur emplacement pour subdiviser les UPE en grappes de la taille requise. Si ces données ne sont pas adéquates, on trace de nouvelles cartes des régions et on les envoie aux BR pour qu'ils procèdent au dénombrement sur le terrain. Le dénombrement sur le terrain consiste à visiter la région et à relever tous les logements qui se trouvent à l'intérieur des limites de l'UPE. Le nombre de logements qui se trouvent entre des caractéristiques physiques permanentes (rivières, routes, etc.) est alors indiqué sur les cartes. Dès que ces cartes sont reçues à la Sous-section du contrôle de l'échantillon, on procède à la formation des grappes.

Formation des grappes

Pour former les grappes, on divise une grande région en de petites régions selon des critères se rapportant aux facteurs de densité optimale et aux fractions de sondage dans la strate. Dans les régions urbaines, les grappes sont des îlots, des groupes d'îlots ou des côtés d'îlot, tandis que dans les régions rurales ce sont généralement des secteurs de dénombrement du recensement. Les SD où le nombre de logements est trop élevé ou trop faible sont, selon le cas, subdivisés ou combinés afin qu'on puisse former le nombre de grappes nécessaires pour représenter les six groupes de renouvellement.

Environ six mois avant le début prévu des interviews, on détermine quelles grappes seront introduites dans l'échantillon des secteurs AR et des secteurs NAR. Celles-ci sont dessinées sur les Formule 01 (schémas de grappe) sur lesquelles des données de contrôle sont préimprimées. Les schémas de grappe sont expédiés aux BR où l'on y joint les formules de listage (formules 02) correspondantes, et le tout est envoyé aux intervieweurs en vue du listage des logements.

Listage

Le listage consiste à attribuer à chaque logement un numéro de ligne de listage qui lui est propre et à représenter ce logement sur la formule 02 par son adresse de voirie ou une description. Tous les logements habitables, qu'ils soient vacants ou occupés, à l'intérieur des limites indiquées sur le schéma de grappe sont inscrits sur la liste. On considère

listed. Listed as a dwelling is any set of living quarters that (i) is structurally separate from the living quarters of other dwellings and (ii) has a private entrance outside the building or a private entrance from a common hall or stairway inside the building. The entrance must be one that can be used without passing through the living quarters of another dwelling.

Collective Dwellings

Structures that are intended to accommodate ten or more unrelated persons are considered collective dwellings. These include religious institutions, homes for the aged, hostels, YWCAs, YMCAs, hospitals, nursing homes, hotels, motels, boarding houses, construction or mining bunk houses, etc. When a collective dwelling is encountered during listing it is temporarily assigned one listing line and details about the dwelling are reported to the Sample Control Unit at Head Office. The Sample Control staff research the Census data available on the collective to determine first, if the dwelling is a component of the Special Area Frame and therefore should be excluded from the area cluster; second, if, based on the type of dwelling, some of the residents are considered inmates and therefore to be excluded from enumeration; and third, the appropriate number of listing lines to allocate to the dwelling to accommodate the non-inmate residents. Collective Dwellings of less than ten residents are automatically allocated one listing line. A dwelling occupied by a single family of ten or more persons would not be considered a collective dwelling.

The period of time allowed for listing serves two functions in the Regional Office. First, it allows flexibility in scheduling listing at the most convenient time to minimize travel and other costs. Second, it allows adjustment to interviewer assignments for changes in the sample location or expected sample take.

Sub-sampling Growth Areas

Because of the systematic selection of dwellings from clusters, a cluster that has experienced considerable growth would have a corresponding increase in expected sample take. This is normally accounted for by adjusting assignments among existing interviewers or, if necessary, by creating new assignments. If these methods prove too costly and the increase is significant, ROs may request that the cluster be sub-sampled. This is done in the Sample Control Unit by either sub-dividing the cluster into several sub-clusters and sampling some of them or by amending the cluster sampling rate to reduce the sample take. In either case a cluster dwelling weight is calculated to compensate for the sub-sampling and this weight is assigned to each dwelling in that cluster during the weighting process.

comme un logement les pièces d'habitation qui : (1) sont structurellement distinctes des pièces d'habitation d'autres logements; (2) ont soit une entrée privée à l'extérieur de l'immeuble soit une entrée privée donnant sur un couloir ou un escalier commun à l'intérieur de l'immeuble. On doit pouvoir utiliser l'entrée sans avoir à traverser les pièces d'habitation d'un autre logement.

Logements collectifs

On considère comme des logements collectifs les constructions (établissements religieux, foyers pour personnes âgées, YWCA, YMCA, hôpitaux, maisons de soins infirmiers, hôtels, motels, maisons de chambres et pensions, camps de chantier où sont les logés les mineurs et les travailleurs de la construction, etc.) prévues pour l'hébergement d'au moins 10 personnes non apparentées. Lorsqu'on tombe sur un logement collectif pendant le listage, on lui attribue temporairement une ligne de listage et on en signale l'existence à la Sous-section du contrôle de l'échantillon au bureau central. Le personnel de la Sous-section consulte les données de recensement sur le logement collectif pour déterminer, premièrement, si ce logement fait partie de la base de secteurs spéciaux et devrait donc être exclu de la base areolaire, deuxièmement s'il s'agit d'un logement dont les occupants sont considérés comme des pensionnaires d'établissement institutionnel et qui devrait donc être exclu du dénombrement et, troisièmement, si le logement collectif doit rester dans la grappe, combien de lignes de listage lui attribuer compte tenu du nombre de résidents. On attribue automatiquement une ligne de listage aux logements collectifs habités par moins de 10 personnes. On ne considère pas comme un logement collectif un logement occupé par une seule famille de 10 personnes ou plus.

La période réservée au listage remplit deux rôles. Premièrement, elle donne suffisamment de jeu pour qu'on puisse organiser cette activité au moment qui permet de réduire au maximum les coûts, de déplacement et autres. Deuxièmement, elle permet de modifier la tâche des intervieweurs si l'emplacement des unités ou le rendement prévu de l'échantillon changent.

Sous-échantillonnage dans les régions en expansion

Comme les logements sont prélevés dans les grappes par la méthode d'échantillonnage systématique, si la population d'une grappe s'accroît considérablement, le rendement prévu de l'échantillon augmente en conséquence. On réagit normalement à cette situation en répartissant les nouveaux logements entre les intervieweurs ou, s'il le faut, en créant de nouvelles tâches. Mais si cette solution revient trop cher et que la croissance de la population est vraiment trop grande, le BR peut demander que la grappe en cause soit sous-échantillonnée. La Sous-section du contrôle de l'échantillon s'en charge, et elle peut soit former des sous-grappes à partir de la grappe et en échantillonner quelques-unes, soit modifier la fraction de sondage pour réduire le rendement de l'échantillon. Dans un cas comme dans l'autre, on calcule un poids ayant pour objet de compenser l'effet du sous-échantillonnage et on l'attribue à chacun des logements de la grappe à l'étape de la pondération.

All listing information is captured on the RO minicomputers and transmitted to head office. ROs are requested to transmit listings approximately 13 weeks before interviewing starts. This allows time for confirmation that the listings have been entered into the main computer and ensures that the information can be verified prior to the selection of dwellings, which occurs six weeks prior to interviewing.

Sample Selection

The Sample Control Unit maintains the sampling programs, which specify the sampling parameters (survey date, survey identification, random start and sampling ratio). These programs are used to select dwellings belonging to the specific rotation group of the particular month. They identify the line numbers representing the dwellings that are to be interviewed. At the same time, the program assigns a unique Docket Number to each selected line number. Additional programs create a control file of all Docket Numbers, which in effect becomes the file for which survey data is expected. The results of the sampling are transmitted back to the Regional Offices and there, new Listing Forms are printed showing all dwellings, and those that have been selected for the forthcoming survey. At the same time, the other survey forms (Household Record Docket F03, Labour Force Questionnaire F05, Rent Schedule F04 and Supplementary Survey Questionnaire F06) are pre-printed with control information.

7.5 Maintenance

Volumes

Prior to the 1981 LFS redesign, there were approximately 12,000 clusters active at any given time in the survey and approximately 900 clusters needed to be replaced each month. With the elimination of a stage of selection and with the increased size of clusters in the NSR in the new design, there are approximately 11,000 active clusters with about 300 needing replacement each month.

Frame

All clusters in SR areas were determined at the time of redesign. Maps covering all SR areas and showing all clusters are maintained in the Sample Control Unit and are used when it is necessary to draw Cluster Diagrams for replacement clusters or to resolve boundary problems that may be reported from the field.

As time passes the counts assigned to clusters become outdated due to differential growth rates and other city development. Although estimates remain unbiased, their variance increases. For this reason, the SR Update Program is maintained between major redesigns. Under this program independent sources are used to identify SR areas where growth or other factors have made counts obsolete. These sources

Tous les renseignements relatifs au listage sont introduits dans les mini-ordinateurs des BR et transmis au bureau central. Les BR doivent transmettre les listes environ 13 semaines avant le début des interviews. Cela laisse le temps de confirmer que les listes ont bien été introduites dans l'ordinateur principal et de vérifier les renseignements avant l'échantillonnage des logements, six semaines avant les interviews.

Prélèvement de l'échantillon

La Sous-section du contrôle de l'échantillon assure la maintenance des programmes d'échantillonnage qui précisent les paramètres (date de l'enquête, identification de l'enquête, origine aléatoire et fraction de sondage) dont on se sert pour choisir les logements dans le groupe de renouvellement utilisé un certain mois. Ces programmes choisissent les numéros de ligne correspondant aux logements où les interviews vont avoir lieu. Ils attribuent en même temps un numéro de dossier à chaque numéro de ligne (logement) choisi. D'autres programmes créent un fichier de contrôle contenant tous les numéros de dossier, c'est-à-dire l'identification des ménages au sujet desquels on s'attend à recueillir des données d'enquête. Le résultat de l'échantillonnage est renvoyé aux BR, où de nouvelles formules de listage sont imprimées indiquant tous les logements et ceux qui ont été échantillonnés pour l'enquête suivante. Les données de contrôle sont également imprimées sur les autres formules d'enquête (Dossier du ménage, F03, Questionnaire de l'enquête sur la population active, F05, Questionnaire sur le loyer, F04, et Questionnaire de l'enquête supplémentaire, F06).

7.5 Mise à jour

Nombre de grappes

Avant le remaniement, il y avait en tout temps environ 12,000 grappes actives dans l'échantillon de l'enquête et environ 900 grappes à remplacer chaque mois. Maintenant qu'un degré d'échantillonnage a été éliminé et que la taille des grappes dans le nouveau plan de sondage des secteurs NAR a été agrandie, il y aura environ 11,000 grappes actives dont 300 devront être remplacées chaque mois.

Base de sondage

Toutes les grappes dans les secteurs AR ont été déterminées au moment du remaniement. Les cartes de ces secteurs, sur lesquelles les grappes sont représentées, sont conservées et mises à jour par le personnel de la Sous-section du contrôle de l'échantillon qui s'en sert pour tracer les schémas des grappes de remplacement ou pour résoudre les problèmes de délimitation signalés par le personnel sur le terrain.

Le nombre de logements dans les grappes devient inexact au fil du temps en raison des écarts de croissance d'une région à une autre et de la façon dont les villes évoluent. Bien que les estimations demeurent non biaisées, leur variance augmente. C'est pour cela qu'on applique le programme de mise à jour des secteurs AR entre les remaniements. Ce programme fait appel à des sources indépendantes, notamment les dossiers administratifs tenus

may be administrative records maintained in municipal offices or data from the quinquennial census. New counts are then obtained for clusters. These counts come from census or from actual field counts. Where necessary, large clusters are sub-divided. New selections are then made within random groups to maximize the retention of originally selected clusters.

Maps for NSR Primary Sampling Units showing all clusters are only available for selected PSUs. As new PSUs are selected, sets of maps are prepared and maintained in the Sample Control Unit.

The separate Apartment Sample Frame is an open-ended list of all apartment buildings in 18 of the largest cities. Each month the Sample Control Unit receives lists of apartment building completions from Central Mortgage and Housing Corporation. For cities that have a separate Apartment frame, Apartment Cards are completed for all apartment projects of 30 or more units. The cards are then sent to the ROs for additional information to confirm the eligibility of the building for inclusion in the Apartment Frame. Eligible buildings are those with 30 or more units and five or more floors of living quarters. These buildings are added to the frame when the completed cards are returned from ROs. In an apartment frame, each apartment building is a separate cluster. Apartment buildings, whether in the apartment frame or located in an area cluster, are listed by allocating one listing line to each unit.

Each Cluster List is in fact open-ended. New dwellings constructed within the boundaries of the cluster are added to the list by the interviewer when he or she visits the cluster. Dwellings are added when they become habitable. Information printed on the Form 02 indicates which of these additions to the list fall into the systematic sample and therefore must be enumerated.

Special Surveys

The LFS Sample Frame and sampling procedures are also used to conduct special surveys, such as the Survey of Consumer Finance, the Food Expenditure Survey and the Family Expenditure Survey. When a special sample is selected, systematic samples represented by random starts that have not yet been used by the LFS are selected from cluster lists. This procedure ensures that the samples are independent of the LFS but are located for the most part within already active clusters. Thus, field costs related to counting and listing are minimized. Listing and counting would only be necessary if there are no more starts within a cluster or PSU that have not been used. The programs, or parallel ones, used for controlling the sample can be used to generate Listing Forms specific to the special survey (a separate code is used to identify dwellings reserved for special surveys). The

par les municipalités et les données tirées du recensement quinquennal, pour repérer les secteurs AR sur lesquels les chiffres ne sont plus bons soit à cause de la croissance de la population soit à cause d'autres facteurs. On obtient alors de nouveaux chiffres de population pour les grappes provenant du recensement ou d'un dénombrement sur le terrain, et, s'il le faut, on subdivise les très grandes grappes. On prélève ensuite un échantillon dans les groupes aléatoires de manière à conserver le plus grand nombre possible de grappes qui avaient été choisies au départ.

On ne dispose de cartes représentant toutes les grappes que pour certaines UPE dans les secteurs NAR. Au fur et à mesure que de nouvelles unités sont échantillonnées, de nouvelles cartes sont tracées et tenues à jour à la Sous-section du contrôle de l'échantillon.

La base de sondage des appartements est une liste ouverte de tous les immeubles d'appartements dans les 18 plus grandes villes du Canada. Chaque mois, la Sous-section du contrôle de l'échantillon reçoit de la Société canadienne d'hypothèques et de logement la liste des immeubles dont la construction est achevée. On remplit une fiche d'appartement pour chacun des immeubles de trente unités ou plus construits dans les villes qui ont leur propre base d'appartements et on envoie cette fiche au BR en vue de recueillir des renseignements supplémentaires qui permettront de confirmer si l'immeuble devrait bien figurer dans la base. Les immeubles admissibles sont les immeubles de cinq étages ou plus qui comptent au moins trente unités; ils sont ajoutés à la base aussitôt que les fiches remplies reviennent des BR. Dans une base d'appartements, chaque immeuble constitue une grappe distincte. Lorsqu'on fait la liste de grappe pour un immeuble d'appartements, que celui-ci fasse partie de la base d'appartements ou de la base aréolaire, on attribue une ligne de liste à chaque unité.

Chaque liste de grappe est ouverte. Quand l'intervieweur visite la grappe, il ajoute à la liste les nouveaux logements qui y ont été construits aussitôt que ceux-ci deviennent habitables. Les renseignements imprimés sur la formule 02 indiquent lesquels des immeubles ajoutés à la liste font partie de l'échantillon systématique et doivent donc participer à l'enquête.

Enquêtes spéciales

On a recours à la base de sondage et aux méthodes d'échantillonnage de l'EPA pour prélever l'échantillon des enquêtes spéciales comme l'enquête sur les finances des consommateurs, l'enquête sur les dépenses alimentaires et l'enquête sur les dépenses des familles. Pour former l'échantillon d'une enquête spéciale, on prélève des échantillons systématiques dans les listes de grappe en partant d'origines choisies au hasard dont on ne s'est pas encore servies pour l'EPA. On s'assure ainsi que les échantillons sont distincts de celui de l'EPA mais que les unités se trouvent pour la plupart dans des grappes actives, ce qui réduit au maximum le coût des activités sur le terrain comme le dénombrement et l'établissement des listes. Ces activités deviennent nécessaires uniquement lorsque toutes les origines aléatoires dans une grappe ou dans une UPE ont déjà été utilisées. On peut se servir des programmes utilisés pour contrôler l'échantillon ou de programmes semblables pour générer les formules de

flexibility in the design also allows sample units to be reserved in clusters or PSUs which the LFS will never use.

7.6 Data Collection and Capture Activities

This section outlines the main steps in the data collection stage of the survey. These functions are carried out by staff in the eight Regional Offices. Each Regional Office is responsible for receiving the computer files from Head Office. These files specify the sample and print the necessary forms, preparing assignment packages, shipping assignments to interviewers, supervising interviewing, receiving completed survey documents, data capture and transmission to Head Office.

Assignment Planning

Five to six weeks before a particular interview week, the RO receives, via the minicomputer network, a cluster file which identifies all the clusters that will be active in the sample. An assignment planning report is then produced. Clusters that have been exhausted are removed and new ones are added.

For each new cluster coming into the survey, regional managers of the LFS identify the assignment number to which the cluster will be allocated, the interviewer number, the language that will be used for printing survey forms, and the printing priority for the assignment. Changes to assignments (such as replacing an interviewer) are also made at this time. New and changed entries to the assignment planning report are transmitted to Head Office no later than four weeks before the survey week in question. This information will be used to control the priority of printing assignments, the correct language of survey forms for the assignment and the batching together of all pre-printed forms for each assignment.

Assignment Preparation

The survey cycle for a given month actually begins when the processing of the data from the previous month ends. Information for each dwelling selected for the month in question is transmitted from Head Office two to three weeks prior to survey week. This allows time for printing the survey forms and sending assignments to interviewers. For the dwellings entering the survey for the first time only control information is transmitted, which is taken from the cluster listings. For the five-sixths of the sample that was in the survey in the previous month, selected data from the Household Record Docket and each questionnaire are also included. With this information, survey forms for the current month are printed (F03, F04, F05, and F06, if applicable).

listage propres à l'enquête spéciale (un code distinct permet de signaler les logements réservés aux enquêtes spéciales). Cette caractéristique du plan de sondage permet aussi de réserver des échantillons dans les grappes ou les UPE qui ne seront jamais utilisées pour l'EPA.

7.6 Collecte et saisie des données

Dans cette section, nous allons décrire les principales étapes de la collecte des données. Les fonctions qui y sont associées sont exécutées par le personnel des huit bureaux régionaux, dont les responsabilités sont les suivantes : recevoir les fichiers informatiques provenant du bureau central qui indiquent quelles unités font partie de l'échantillon et permettent d'imprimer les formules nécessaires, préparer les documents de la tâche, envoyer ces documents aux intervieweurs, superviser les interviews, recevoir les documents remplis, saisir les données et transmettre celles-ci au bureau central.

Planification des tâches

Cinq ou six semaines avant la semaine où les interviews doivent avoir lieu, le BR reçoit, par le réseau de mini-ordinateurs, un fichier de grappes indiquant toutes les grappes qui seront actives dans l'échantillon, et un rapport de planification de la tâche est produit. Les grappes épuisées auront été retirées de ce fichier et les nouvelles grappes y auront été ajoutées.

Les chefs régionaux de l'EPA indiquent les renseignements suivants au sujet de chaque nouvelle grappe introduite dans l'échantillon : le numéro de la tâche dont chaque grappe fera partie, le numéro de l'intervieweur, la langue dans laquelle les formules d'enquête seront imprimées et l'ordre d'impression des documents de la tâche. C'est à cette étape aussi que sont effectués les changements concernant les tâches comme le remplacement des intervieweurs. Les ajouts et les modifications apportés au rapport de planification de la tâche sont communiqués au bureau central dans les quatre semaines précédant la semaine d'enquête. On se servira de ces renseignements pour vérifier si l'ordre d'impression des documents de la tâche est respecté, si les formules d'enquête sont dans la bonne langue et si toutes les formules pré-imprimées relatives à une tâche ont été correctement mises en lot.

Préparation de la tâche

Le cycle de l'enquête pour un mois donné commence aussitôt que s'achève le traitement des données relatives au mois précédent. Les renseignements concernant chaque logement dans l'échantillon pour le mois en cours sont transmis au bureau central deux ou trois semaines avant la semaine d'enquête, pour qu'on y ait le temps d'imprimer les formules d'enquête et d'envoyer leur tâche aux intervieweurs. Pour ce qui est des logements qui participent à l'enquête pour la première fois, il s'agit des données de contrôle seulement, provenant des listes de grappe. Pour les cinq sixièmes de l'échantillon qui participaient à l'enquête le mois précédent, on inclut également certaines données tirées du dossier du ménage et du questionnaire. On se sert de cette information pour imprimer les formules d'enquête pour le mois en cours (F03, F04, F05, et F06 s'il y a lieu).

In large cities, where listing information is adequate, postal addresses are taken from cluster lists and used to mail out introductory letters to dwellings selected for the first time. Where this is not possible, interviewers have a supply of letters which can be used to introduce the survey at the first contact with the household.

In the second week prior to survey week, clerical staff in the ROs prepare and ship assignment packages to interviewers. Packages include pre-printed survey forms, blank survey forms, letters of introduction, LFS brochures, maps and listings, and control and expense claim forms.

Collection Methodology

Interviews are conducted Monday to Saturday of survey week. An interviewer's assignment will typically contain 60 dwellings. First interviews are normally done by face-to-face contact at the dwelling; all subsequent interviews may be conducted by telephone. To minimize the number of times an interviewer must visit a dwelling in order to obtain the first interview, a back-up procedure using telephones for first contacts has been developed. When several visits to a dwelling has not resulted in contacting an eligible respondent, attempts are made to obtain telephone numbers from sources such as city directories and neighbours. Dwellings are then telephoned either to arrange for a personal visit or to conduct the first interview by telephone. During July and August, months when temporary absence from dwellings is high, the interview period may be extended to Sunday or Monday.

Survey information is obtained directly, if possible, from all household members present, although any household member 15 years of age or more may be interviewed to obtain proxy responses for all other household members.

Household membership consists of all persons living in the selected dwelling who have no usual place of residence elsewhere. Members of the Armed Forces are included in the household listing although they are excluded from the survey. Visitors from other countries and representatives of foreign governments, as well as inmates of institutions, are not included.

Receipt and Capture

Interviewers are instructed to return all completed survey forms to the ROs daily. Theoretically, all questionnaires should have been returned by the Saturday of survey week, although in practice this is not done until the Tuesday or Wednesday of the following week.

The ROs begin to process the received forms on the Tuesday of survey week. Clerical staff first perform a set of basic edits (for example, they ensure that a response code is entered on each F03 and that there is a completed F05 for each eligible household member).

Dans les grandes villes, si les renseignements inscrits sur les listes sont adéquats, on se sert des adresses de voirie qui s'y trouvent pour envoyer une lettre de présentation aux logements échantillonnés pour la première fois. Sinon, les intervieweurs ont une réserve de lettres qu'ils peuvent utiliser pour présenter l'enquête lors du premier contact avec le ménage.

Au cours de la deuxième semaine précédant la semaine d'enquête, le personnel de bureau aux BR prépare et expédie aux intervieweurs les documents de la tâche. Ces documents comprennent : les formules d'enquête préimprimées, les formules d'enquête vierges, les lettres de présentation, les brochures de l'EPA, les cartes et les listes et enfin les formules de contrôle et de demande de remboursement des frais.

Méthode de collecte

Les interviews ont lieu du lundi au samedi de la semaine d'enquête. En général, la tâche d'un intervieweur compte 60 logements. Normalement, la première interview est faite en personne au logement et les suivantes, par téléphone. Pour réduire au maximum le nombre de visites effectuées au logement en vue d'obtenir la première interview et le nombre de contacts manqués, on a mis au point une marche à suivre selon laquelle on se sert du téléphone en vue d'établir le premier contact. Lorsqu'après un certain nombre de visites au logement l'intervieweur n'a pas réussi à prendre contact avec un membre admissible du ménage, il doit essayer d'obtenir le numéro de téléphone du ménage soit en consultant l'annuaire, soit en le demandant à un voisin, et téléphoner pour prendre rendez-vous pour une interview sur place ou pour effectuer directement la première interview. Pendant les mois de juillet et d'août, étant donné que le nombre de ménages temporairement absents de leur logement est élevé, les intervieweurs peuvent prolonger la période d'interview jusqu'au dimanche ou lundi.

Si c'est possible, les intervieweurs recueillent les données d'enquête directement auprès de tous les membres du ménage présents, mais ils peuvent aussi interviewer n'importe quel membre du ménage âgé de 15 ans ou plus et obtenir des réponses par procuration au sujet des autres membres du ménage.

Les membres d'un ménage sont toutes les personnes qui habitent un logement échantillonné et n'ont pas d'autre domicile habituel. Les membres des forces armées sont inclus dans la liste des ménages bien qu'ils soient exclus de l'enquête. Les étrangers en visite au Canada et ceux qui représentent un autre Etat de même que les pensionnaire d'un établissement institutionnel sont également exclus.

Réception des formules et saisie des données

Les intervieweurs doivent retourner toutes les formules d'enquête remplies aux BR au jour le jour. En principe, tous les questionnaires devraient être partis quand arrive le samedi de la semaine d'enquête, mais en pratique ils ne le sont pas avant le mardi ou le mercredi de la semaine suivante.

Le dépouillement des formules reçues aux BR débute le mardi de la semaine d'enquête. Le personnel de bureau commence par faire une série de vérifications de base (par exemple, s'assurer qu'il y a un code de réponse sur chaque F03 et qu'il y a une F05 remplie pour chaque membre

Edited forms are separated into "births" and "non-births" and bundled into batches of approximately 25 F03s and their related F04s, F05s and F06s. Batches are sent to the Computer Operations unit, where the information is captured on the minicomputer. Transmission of the captured data to Head Office takes place nightly, beginning Tuesday of survey week. The cut-off for the receipt of questionnaires from interviewers is the Wednesday noon after survey week. The processing of these questionnaires takes place Wednesday afternoon, with the final transmission to Ottawa that night.

Verification of Data Capture

To monitor and control the quality of data capture, prior to the transmission of batches of captured survey forms, they are subject to verification on either a 100 percent basis or a sample basis. Batches of F03s for birth households or F03s that have changes are verified by having all forms data-captured a second time, matched and reconciled. For other survey forms, a sample of documents from each batch are re-captured and matched. All identified errors are corrected. If the number of errors exceeds a cut-off value, the whole batch is re-captured. The amount of sample verification varies according to the type of survey form and the past performance of the data entry operators.

7.7 Data Processing

This section briefly describes the major steps in the processing and dissemination of the monthly Labour Force Survey data in Head Office. The Labour Force Survey data are released each month on the second Friday after survey week. In order to meet this deadline a very tight schedule is followed in Head Office. In fact, the survey cycle begins at the end of the previous month's survey when the information used in preprinting the F03s and F05s is transmitted to the Regional Offices. As detailed earlier, the forms are preprinted, the interviewer's assignments sent out, and, in survey week, the interviewing takes place. By Wednesday of survey week transmission of the data to Head Office for processing begins. The file used to receive the transmitted data, called the structure and response file, contains the F03 data for each household and the F05 data for each member of the household. As the data are received they are edited and coded. By Wednesday of the week following survey week, all the survey data has been received from the Regional Offices and by Friday of that same week a file is ready for weighting. The weighted file is used to produce the tabulations for the monthly press release and publication.

admissible du ménage). Les formules vérifiées sont ensuite séparées selon qu'elles se rapportent à de nouveaux ou à d'anciens ménages et mises en lots d'environ 25 F03 accompagnées des F04, F05 et F06 correspondantes. Les lots sont envoyés à la Sous-section des travaux informatiques où les données sont introduites dans le mini-ordinateur. La transmission au bureau central des données saisies se fait chaque nuit à partir du mardi de la semaine d'enquête. La date limite pour la réception des questionnaires envoyés par les intervieweurs est le mercredi midi suivant la semaine d'enquête. Le dépouillement de ces questionnaires a lieu le mercredi après-midi et la dernière transmission des données à Ottawa, dans la nuit qui suit.

Vérification de la saisie des données

Pour surveiller et contrôler la qualité de la saisie des données, après cette opération, on soumet les lots de formules à une vérification intégrale ou par échantillon. Dans le cas des F03 se rapportant à de nouveaux ménages ou ayant subi certains changements, on en prend des lots complets et on saisit toutes les données une seconde fois, après quoi on compare les résultats obtenus avec les premiers. Dans le cas des autres formules d'enquête, on en prend un échantillon, on saisit les données de nouveaux et on fait la comparaison. Toutes les erreurs relevées sont corrigées. Si le nombre d'erreurs dépasse un certain seuil, le lot complet est soumis à la saisie de nouveau. La taille de l'échantillon soumis à la vérification varie selon le genre de formule et le rendement passé des préposés à l'introduction des données.

7.7 Traitement des données

Dans cette section, nous allons décrire brièvement les principales étapes du traitement et de la diffusion des données mensuelles de l'enquête sur la population active au bureau central. Les données tirées de l'enquête sur la population active sont diffusées chaque mois le deuxième vendredi suivant la semaine d'enquête. Pour arriver à respecter cette échéance, il faut suivre un calendrier très serré au bureau central. De fait, le cycle de l'enquête commence à la fin de l'enquête du mois précédent avec la transmission aux BR des données nécessaires à la préimpression des F03 et des F05. Comme nous l'avons vu plus haut, les formules sont préimprimées, la tâche des intervieweurs est expédiée puis les interviews ont lieu (pendant la semaine d'enquête). La transmission des données au bureau central en vue de leur traitement débute le mercredi de la semaine d'enquête. Le fichier dans lequel sont versées les données reçues, qu'on appelle le fichier structuré des réponses, contient les données sur chaque ménage provenant de la formule F03 et les données sur chaque membre des ménages provenant de la formule F05. Les données sont vérifiées et codées au fur et à mesure qu'elles arrivent. Le mercredi de la semaine suivant la semaine d'enquête, toutes les données d'enquête ont été reçues des bureaux régionaux et le vendredi, le fichier est prêt pour la pondération. On se sert du fichier pondéré pour faire les totalisations qui seront reproduites dans la publication et le communiqué mensuels.

Editing and Coding

Producing a file for weighting from the structure and response file involves five processing steps, each one using a combination of computerized and manual procedures. These are Record Acceptance, Notes Coding, Industry/Occupation Coding, Detail Edits and Demographic Edits. Before going through these steps, the F03 and F05 data are reformatted onto the consolidated file. Each record on this file contains the F05 data for the respondent, plus the F03 data for the household containing the respondent.

Record Acceptance

The function of this process is to ensure F05 information exists for each eligible person aged 15 and over in the household. Conversely, it also ensures that anyone who should not have any F05 data in fact does not (i.e., respondents under 15 years and military personnel). If a record fails either of these checks, it is output for correction by a member of the processing staff. After examining the information given for the record in error and the responses of the other household members, action is taken to correct the error. For example, if a respondent 15 years and over has no F05 data and the rest of the household data indicates that the person should have an F05, the data will be imputed using the F05 data of a respondent in the previous month's survey who has similar demographic data from the same geographic area. If, on the other hand, the rest of the household data indicates the person is no longer a member of the household, the data on this person will be deleted from the file.

Notes Coding

In this process the notes for certain key questions entered by the interviewer are examined by the processing staff when the response to the question is "Other". If a response other than "Other" should have been entered the response will be changed; otherwise it will remain as is.

Industry/Occupation Coding

In this process, industry and occupation codes are assigned using the respondent's job description on the F05. In order to decrease the volume of coding each month, only new respondents to the survey, or respondents whose job description has changed since the previous month go through the industry occupation coding procedures. If the job description is unchanged from the previous month, the codes from the last month are simply brought forward onto the current month's data automatically.

For records that do require industry occupation coding, the first step is to attempt to code the record with a computerized procedure. If this is unsuccessful, the record will be output for manual coding.

Vérification et codage

Pour produire un fichier en vue de la pondération à partir du fichier structuré des réponses, il faut suivre cinq étapes comprenant des opérations informatiques et des opérations manuelles : l'acceptation des dossiers, le codage des notes, le codage de la branche d'activité et de la profession, la vérification des détails et la vérification des données démographiques. Avant le commencement de ces opérations, les données tirées des formules F03 et F05 sont combinées et versées dans le fichier intégré. Chaque enregistrement dans ce fichier contient les données sur le répondant provenant de la formule F05 et celles, provenant de la formule F03, sur le ménage dont fait partie le répondant.

Acceptation des dossiers

L'objet de cette étape est de s'assurer qu'on dispose bien de données tirées de la formule F05 sur chaque personne admissible dans le ménage âgée de 15 ans et plus et, inversement, qu'aucune donnée au sujet des personnes inadmissible est présente (c'est-à-dire sur les personnes âgées de moins de 15 ans et sur le personnel militaire). Si un enregistrement est rejeté à une de ces deux vérifications, il est imprimé afin qu'un préposé au traitement y apporte les corrections nécessaires. Le préposé examine les renseignements que contient l'enregistrement erroné et les réponses fournies par les autres membres du ménage, puis il prend une mesure corrective. Si, par exemple, il n'y a pas de données provenant d'une formule F05 sur un répondant âgé de 15 ans ou plus et que les données sur le ménage indiquent qu'il devrait y en avoir, on comble cette lacune par imputation, c'est-à-dire en prenant des données tirées d'une formule F05 sur un répondant qui a participé à l'enquête le mois précédent, dont les caractéristiques démographiques sont semblables et qui demeure dans la même région. Si, au contraire, on a des données sur une personne qui, selon les données sur le ménage, ne fait plus partie de celui-ci, on les supprime dans le fichier.

Codage des notes

Cette étape consiste, pour le personnel chargé du traitement des données, à examiner le contenu des notes, c'est-à-dire de la section où l'intervieweur a inscrit des observations vis-à-vis de questions importantes auxquelles la réponse est "autre". S'il avait fallu choisir une réponse différente, on la change, sinon on la laisse telle quelle.

Codage de la branche d'activité et de la profession

Cette étape consiste à coder la branche d'activité et la profession en fonction de la description du travail effectué par le répondant que celui-ci a fournie et qui est inscrite sur la formule 05. Pour réduire le nombre de réponses à coder chaque mois, on ne procède au codage de la branche d'activité et de la profession que dans le cas des nouveaux répondants ou des répondants dont la réponse a changé par rapport au mois précédent. Si la réponse n'a pas changé, on se contente de rappeler le code du mois précédent et de l'insérer automatiquement dans les données du mois en cours.

En ce qui concerne les enregistrements qui doivent être codés, on commence par essayer une procédure automatique. Si elle ne fonctionne pas, l'enregistrement est imprimé et on procède au codage à la main.

The automatic coding system uses a library of previous coded descriptions. Initially, an attempt is made to match the description requiring coding to this library of coded descriptions. This saves time so that frequently occurring descriptions do not need to be coded in the more complicated second phase of the automatic coding. This second phase is based on the Hellerman Algorithm developed at the U.S. Bureau of the Census (Hellerman, 1981). This algorithm will assign a code to a description using a matching coded description. If no match is available, it will assign the code using a description similar to that being coded. A similar description would be one that differs from the description being coded by only one or two words.

If the automatic coding system is not successful in assigning the industry and occupation codes, the record will be output to be coded manually. The automatic system is designed to allow the partial coding of the job description. Therefore, a record output for manual coding may only need to be assigned an industry code or an occupation code. The Standard Industrial Classifications and Standard Occupation Classifications manuals are used in manual coding.

Once the industry and occupation codes have been assigned, a logic check is done to ensure that the combination of industry, occupation and class of work codes "make sense". If the codes that were assigned to the job description are illogical, the record will be output for coding again.

Detail Edits

This is the edit and imputation process for the labour force information on the F05. Here, validity and logic errors are detected and corrected. Generally, any invalid or illogical responses are automatically corrected either deterministically, using the previous month's data, or by "cold-deck" imputation. Deterministic corrections are those in which the correct response can be determined by the responses to other questions on the form. The previous month's response to a question in error in the current month is used to correct the current month's data if the responses to the other questions are the same in both months.

In cold-deck imputation, the response given the previous month on another record in the same geographic area and with similar characteristics as the record in error is used in place of the response to the question in error. If neither of these two methods can be used to correct this error, the record is printed and corrected manually. The processing staff have instructions on how to correct the error either by looking at the responses given this month or the previous month, or by imputing a correct answer from a record with similar characteristics in the previous month's data stored on micro fiche.

Le système de codage automatique fait appel à une réserve de descriptions déjà codées. Il commence par comparer la description devant être codée à des descriptions souvent données utilisées et précodées. Cette étape vise à faire gagner du temps en évitant que les descriptions courantes ne soient pas soumises à la deuxième étape du codage automatique, qui est plus compliquée. Cette deuxième étape est basée sur l'algorithme de Hellerman, mis au point au U.S. Bureau of the Census. Cet algorithme permet non seulement d'attribuer un code à une description en trouvant une description pareille déjà codée, mais aussi, si on ne trouve pas de description pareille, de prendre le code d'une description semblable. Une description semblable en est une qui diffère de la description à coder par un ou deux mots.

Si le système de codage automatique ne réussit pas à attribuer un code de branche d'activité ou de profession, l'enregistrement est imprimé et subit alors un codage manuel. Le système automatique peut faire un codage partiel à partir de la description du travail effectué, de sorte qu'il peut ne manquer à un enregistrement rejeté du système que l'un ou l'autre des codes, c'est-à-dire soit celui de la branche d'activité, soit celui de la profession. Les ouvrages de référence pour le codage manuel sont la Classification type des industries et la Classification type des professions.

Une fois que le code de branche d'activité et le code de profession ont été attribués, on fait un contrôle de cohérence pour s'assurer que les codes de branche d'activité, de profession et de catégorie de travailleur vont bien ensemble. Si les codes attribués en fonction de la description du travail effectué ne sont pas plausibles, l'enregistrement est imprimé en vue d'être recodé.

Vérification des détails

Il s'agit du processus de contrôle et d'imputation appliqué aux données relatives à l'activité tirées de la formule F05. C'est à cette étape que les données non valides et les erreurs de logique sont corrigées. Dans la plupart des cas, les réponses non valides et illogiques sont corrigées automatiquement soit de manière déterministe, soit d'après les réponses données le mois précédent, soit par la méthode d'imputation dite du "cold deck". La méthode déterministe est celle qui consiste à déterminer la bonne réponse d'après les réponses données aux autres questions sur la même formule. La deuxième méthode consiste à aller chercher dans l'enregistrement du mois précédent la réponse qui permettra de corriger la réponse erronée dans l'enregistrement du mois en cours, à condition que les réponses aux autres questions soient les mêmes dans les enregistrements des deux mois.

Enfin, selon la méthode du "cold deck", on remplace la réponse erronée par la réponse donnée le mois précédent par une personne de la même région ayant des caractéristiques semblables. Si toutes ces méthodes échouent, l'enregistrement est imprimé et corrigé à la main. Le personnel chargé du traitement des données a des directives sur la façon de corriger l'erreur soit en consultant les autres réponses données le même mois ou le mois précédent, soit en allant chercher dans les microfiches la réponse donnée le mois dernier par une personne ayant des caractéristiques semblables.

Demographic Edits

In this process the demographic data reported on the F03 for each household member are edited in three stages. First, staff ensure that all the responses are valid (i.e., none are out of range or blank), second, they check that the responses are logical at the respondent level and third, they make sure the demographic responses are logical at the household level. An example of a logic error at the individual level would be if the spouse had a marital status of single rather than married. A logic error at the household level would be if there were two spouses in one family. Any errors are printed for manual correction by the processing staff who have instructions on the various methods of correction, depending on the type of error.

Weighing

When the demographic characteristics have been verified, the consolidated file of F03 data for each household and the F05 data for each member of the household are ready to be weighted. In this phase, each file is weighted to indicate the number of persons represented by each individual in the sample, thereby creating a weighted file for purposes of data compilation. See Chapter 8 for details.

Concepts and Compilation

The Labour Force Survey is aimed at providing data on persons of working age. The data is divided into three categories: the employed, the unemployed and persons not in the labour force. A correspondence between the responses of each person in the sample and the concepts represented by the categories can be established according to a set of decision rules. When this phase is complete, the cumulative weights of all the files having the characteristics in question yield detailed estimates on the labour market.

7.8 Dissemination of Data

Statistics from the LFS reflect the state of the labour market at a specific time, namely, the week containing the 15th day of the month. This week is known as the reference week. During the third week after the reference week (usually the first full week of the next month), the estimates are analysed and the main tables are prepared for the production of the publication *Labour Force Information* (catalogue no. 71-001P), which is available to the public at 7 a.m. on the Friday. The complete list of all the tables published regularly in *The Labour Force* (catalogue no. 71-001) is available ten days later.

Publications

To ensure the rapid dissemination of the most recent labour market statistics, only a short commentary is given in *Labour Force Information* (71-001P). This commentary consists of a concise analysis of data from the current survey, the main seasonally adjusted

Vérification des données démographiques

Les données démographiques se rapportant à chaque membre du ménage qui sont inscrites dans la formule F03 sont contrôlées en trois étapes. Premièrement, il faut s'assurer que toutes les réponses sont valides (c'est-à-dire qu'aucune n'est en blanc ni en dehors de la fourchette de valeurs acceptables), deuxièmement, il faut s'assurer qu'elles sont cohérentes au niveau du répondant et, troisièmement, il faut s'assurer qu'elles sont cohérentes au niveau du ménage. Un exemple d'erreur de logique au niveau du répondant serait le cas d'un conjoint dont l'état matrimonial serait "célibataire" au lieu de "marié". Un exemple au niveau du ménage serait une famille où il y aurait deux conjoints. Aussitôt que le personnel chargé du traitement relève une erreur, il imprime l'enregistrement pour prendre les mesures correctives nécessaires selon le genre d'erreur en cause.

Pondération

Lorsque la vérification des caractéristiques est complétée, le fichier consolidé des renseignements des dossiers des ménages (F03) et des individus (F05) est prêt pour la pondération. Cette étape consiste à attribuer à chaque dossier un poids qui désigne le nombre de personnes que représente chaque individu faisant partie de l'échantillon. Au terme de ce procédé, un fichier pondéré sera créé pour fins de compilation des données. Voir le prochain chapitre pour les détails.

Concepts et compilation

L'enquête sur la population active vise à fournir des données pour la population d'âge actif qui est divisée en trois catégories: les personnes occupées, les chômeurs et les inactifs. Un ensemble de règles de décisions permet d'établir la correspondance entre les réponses fournies par chaque personne faisant partie de l'échantillon et les concepts qui définissent ces trois catégories. Ce processus terminé, il est possible de produire des estimations détaillées sur le marché de travail en cumulant les poids de tous les dossiers qui possèdent des caractéristiques données.

7.8 Diffusion des données

Les statistiques de l'enquête sur la population active réfèrent à l'état du marché du travail à un moment précis dans le temps, soit la semaine comprenant le 15 du mois, i.e., la semaine de référence. La troisième semaine suivant la semaine de référence (habituellement la première semaine complète du mois suivant) les estimations sont analysées et les tableaux principaux sont préparés afin de produire la publication *Information population active* (71-001P au catalogue) qui est disponible au public dès 7h00 le vendredi. Le catalogue complet des tableaux régulièrement publiés dans *La population active* (71-001 au catalogue) est disponible dix jours plus tard.

Publications

Afin de permettre la diffusion rapide des plus récentes statistiques sur l'état du marché du travail, la publication *Information population active* (71-001P) ne contient qu'un commentaire, qui analyse brièvement les données courantes de l'enquête, les principales estimations désaisonnalisées et

and unadjusted estimates for Canada and the provinces, as well as unadjusted estimates for metropolitan areas and economic regions. From May to September, data on student activities is included in the publication.

All these data are provided in *The Labour Force (71-001)*, a monthly publication incorporating more than 40 statistical tables giving details of the population's characteristics according to such variables as sex, marital status, age, level of schooling, school attendance, industry and occupation, class of worker, number of hours worked, reasons for part-time employment, duration of unemployment, type of work sought, reasons for leaving last job, family status and reasons for not looking for work. The December issues of *The Labour Force* contain the annual averages of all the monthly tables.

Users of labour force statistics have access to more than 6,000 historical data series in the CANSIM database. These series are updated monthly and are available the day the latest data are released.

Special Compilations

Users requiring more detailed statistics than those available in the publications described above can contact the Information and Current Analysis Section of the Labour Force Survey Sub-division, which provides various services in return for the cost of producing special compilations.

Microdata Tape

Users who wish to produce their own tables, manipulate data and conduct statistical experiments can obtain a microdata tape of the LFS. The 56 survey variables on the tape are selected to ensure the data supplied by respondents remain confidential.

non désaisonnalisées pour le Canada et les provinces, ainsi que les estimations non désaisonnalisées pour les régions métropolitaines et les régions économiques. Pour les mois de mai à septembre inclusivement, les données sur l'activité des étudiants sont ajoutées à cette publication.

Toutes ces informations sont reproduites dans le catalogue mensuel *La population active (71-001)* qui contient 40 tableaux de statistiques détaillant les caractéristiques de la population selon diverses variables telles: le sexe, l'état matrimonial, l'âge, le niveau de scolarité, la fréquentation scolaire, la branche d'activité, la profession, la catégorie de travailleurs, le nombre d'heures travaillées, les raisons données pour l'emploi à temps partiel, la durée du chômage, le genre d'emploi recherché, les raisons données pour avoir quitté le dernier emploi, la situation des familles, les raisons pour ne pas avoir cherché d'emploi. Enfin, les numéros de décembre de *La population active* contiennent les moyennes annuelles pour tous les tableaux publiés mensuellement.

De plus, par l'intermédiaire des bases de données de CANSIM, les utilisateurs de statistiques sur la population active ont accès à plus de 6000 séries chronologiques qui sont mises à jour mensuellement et sont disponibles dès le jour de diffusion des données les plus récentes.

Compilations spéciales

Pour les usagers qui ont besoin de statistiques encore plus détaillées que celles qui sont disponibles dans les publications décrites ci-haut, la section d'information et d'analyse courante de l'Enquête sur la population active offre divers services, moyennant le recouvrement des frais pour la production de compilations spéciales.

Bandes de micro-données

Les usagers qui désirent créer eux-même leurs propres tableaux spéciaux, manipuler les données et faire des expériences statistiques, peuvent se procurer une bande de micro-données de l'enquête. Les 56 variables sur cette bande regroupent les données de façon à assurer la confidentialité des informations fournis par les enquêté-e-s.

Chapter 8

Weighting and Estimation

8.1 Introduction

The principle behind the estimation procedure in a probability survey such as the Labour Force Survey is that each person in the sample "represents" not only himself or herself, but several persons not in the sample as well. For example, in a simple random sample of 2 percent of a population, each person in the sample represents 50 people in the population, since one person out of fifty was selected at random. Given the sample data in the form of records, one way to produce estimates for the population would be to duplicate each record so that there are 50 copies of each one, and then to compile any aggregates and cross-classifications of interest for this "pseudo-population". If the 2 percent sample is truly representative of the real population, then the numbers, i.e., estimates, obtained by this procedure will be close to the values one would obtain if the records for the real population were available.

In a probability sample, the sample design itself determines weights that may be used to produce unbiased estimates. Each person's record may be weighted by the inverse of the probability of selecting that person in the sample. Thus, in the above example, instead of creating a pseudo-population file of duplicate records, one can simply attach a weight of $1/0.02$, or 50, to each sampled person's responses to produce simple estimates.

For the LFS, the file created for tabulation contains one record per person in the sample. Each record contains labour force and demographic characteristics for the respondent. Instead of physically duplicating the sample records, a weight is entered on each record. This weight can be thought of as the number of times the record should be duplicated. If a characteristic is to be estimated, then an estimate can be obtained as the sum of the weights of all persons in the sample with that characteristic. For example, if the characteristic of interest is the number of persons employed in manufacturing, then an estimate of this can be obtained by summing the weights on the records of all persons employed in manufacturing.

Frequently, objective information on relevant characteristics for the population is available from sources other than the survey itself. Several estimation methods exist that use such auxiliary information to increase the reliability of estimates. In the past, the LFS has used ratio estimation and raking ratio estimation to increase reliability. But these methods yield a different final weight for individuals in the same household, resulting in inconsistent estimates for

Chapitre 8

Pondération et estimation

8.1 Introduction

Le principe sur lequel repose la méthode d'estimation utilisée dans une enquête à échantillonnage probabiliste, comme l'enquête sur la population active, est que chaque personne dans l'échantillon "représente" non seulement elle-même, mais aussi plusieurs personnes qui ne sont pas échantillonnées. Par exemple, dans un échantillon aléatoire simple de 2% de la population, chaque personne dans l'échantillon représente 50 personnes dans la population, puisqu'on a choisi au hasard une personne sur cinquante. Comme les données de l'enquête sont inscrites dans un enregistrement, on pourrait, pour produire des estimations relative à la population entière, faire cinquante copies de chaque enregistrement puis calculer les totaux et faire les recoupements qui nous intéressent relativement à cette pseudo-population. Si l'échantillon de 2% est vraiment représentatif de la population réelle, les nombres, c'est-à-dire les estimations, obtenus de cette manière seront voisins des valeurs qu'on obtiendrait si on avait des enregistrements sur toute la population.

Dans un échantillon probabiliste, c'est le plan de sondage qui détermine les poids à utiliser pour produire des estimations non biaisées. On peut notamment attribuer à l'enregistrement se rapportant à chaque personne un poids qui est l'inverse de la probabilité de sélection de cette personne. Ainsi, dans l'exemple qui précède, au lieu de créer un fichier sur une pseudo-population qui serait constituée des enregistrements copiés cinquante fois chacun, on pourrait simplement rattacher aux réponses de chaque personne échantillonnée un poids de $1/0.02$ ou de 50 et produire de la sorte des estimations simples.

Dans l'EPA, le fichier créé aux fins de la totalisation des données compte un enregistrement par personne échantillonnée. Chaque enregistrement contient des données sur les caractéristiques du répondant du point de vue démographique et du point de vue de l'activité. Au lieu de copier les enregistrements, on y inscrit un poids. Ce poids représente le nombre de fois que l'enregistrement devrait être copié. Si on veut produire une estimation relativement à une caractéristique, on peut le faire en additionnant le poids de toutes les personnes dans l'échantillon qui ont cette caractéristique. Si la caractéristique qui nous intéresse est le nombre de personnes qui ont un emploi dans le secteur manufacturier, on peut en obtenir une estimation en additionnant le poids attribué à toutes les personnes travaillant dans ce secteur.

Il arrive souvent qu'on ait accès à des données objectives sur des caractéristiques pertinentes de la population par d'autres sources que l'enquête. Plusieurs méthodes d'estimation se servent de telles données auxiliaires pour augmenter la fiabilité des estimations. Par le passé, on a eu recours à l'estimation par le quotient et à l'estimation itérative du quotient pour augmenter la fiabilité des données de l'EPA. Toutefois, on n'obtient pas par ces méthodes un poids final pareil pour tous les membres d'un même ménage, de sorte

individuals and families. To ensure that all household members have the same final weight, the LFS now uses a method based on a linear regression approach. Like the previous methods, the current one uses census projections for the sizes of various age-sex groups, economic regions and major cities as auxiliary variables. Because of its flexibility, the current method uses a greater number of such variables than before.

In the LFS, the **final weight** attached to each record (i.e., individual) is derived in six steps. The first five steps are used to calculate a **subweight**, which is the product of five factors: the basic weight, the rural-urban factor, a balancing factor for nonresponse, the cluster subweight and the sample stabilization factor. The sixth and final step transforms the subweight into a final weight using the regression-based method mentioned in the previous paragraph. Each of these steps is described in more detail in the following sections.

8.2 The Basic Weight

The basic weight is determined by the sample design of the survey. It is simply the inverse of the sampling ratio. Thus, using the example given in the introduction to this chapter, a 2 percent sample would have a basic weight of $1/0.02$, or 50. The sampling ratio is discussed in Chapter 2.

The basic weight differs among economic regions. Within an economic region it is the same for all Non-Self Representing strata but varies among Self Representing strata.

8.3 Rural-Urban Factor

NSR areas contain rural and urban parts. Information concerning the total population in rural and urban areas is available from the 1981 Census for each Primary Sampling Unit, including those that are not in the sample. Using the sampled PSUs only, a simple estimate of the 1981 rural and urban populations in each province can be obtained by dividing each sampled PSU's 1981 rural and urban populations by the PSU's known probability of selection and summing to the Economic Region level. These Economic Region estimates of the urban and rural populations are then compared with the actual urban and rural populations to judge if the sample is representative. The ratio of actual rural population to estimated rural population is called a rural factor, and an urban factor is defined similarly. These two factors are computed when the PSUs are selected, and are entered on each sample record in the NSR areas. Changes in these factors are incorporated at the time of PSU rotation.

que les estimations relatives aux particuliers et celles relatives aux familles ne sont pas cohérentes. Pour s'assurer que tous les membres d'un ménage ont le même poids final, on utilise à présent une méthode fondée sur la régression linéaire. Comme les méthodes utilisées antérieurement, celle-ci se sert de projections du recensement concernant la taille de divers groupes d'âge-sexe, les régions économiques et les grandes villes en tant que variables auxiliaires. La méthode actuelle étant plus souple, elle a recours à un plus grand nombre de variables de ce genre que l'ancienne.

Dans l'EPA, on calcule en six étapes le **poids final** attribué à chaque enregistrement (et donc à chaque personne). Les cinq premières étapes permettent de calculer un **sous-poids**, qui est le produit de cinq facteurs : le poids de base, le facteur rural-urbain, le facteur de compensation pour la non-réponse, le sous-poids de grappe et le facteur de stabilisation de l'échantillon.

8.2 Poids de base

Le poids de base est déterminé par le plan de sondage de l'enquête. C'est simplement l'inverse de la fraction de sondage. Si l'on reprend donc l'exemple donné au début de ce chapitre, pour un échantillon de 2% de la population, le poids de base serait de $1/0.02$ ou 50. La fraction de sondage est traitée au chapitre 2.

Le poids de base varie selon la région économique. Dans une région économique, il est le même pour toutes les strates NAR mais varie selon la strate AR.

8.3 Facteur rural-urbain

Les secteurs NAR contiennent des parties rurales et des parties urbaines. On connaît la population totale des régions rurales et urbaines de chaque UPE par les chiffres du recensement de 1981, y compris celle des régions qui ne sont pas dans l'échantillon. En se servant seulement des chiffres sur la population des UPE échantillonnées, on peut obtenir une estimation simple de la population rurale et urbaine de chaque province selon les chiffres de 1981 en divisant la population rurale et urbaine de chaque UPE échantillonnée par la probabilité de sélection connue de l'UPE et en additionnant les résultats obtenus jusqu'au niveau de la région économique. On compare alors ces estimations de la population rurale et urbaine des RE aux chiffres réels sur ces populations afin de déterminer si l'échantillon est représentatif. Le rapport entre la population rurale réelle et la population rurale estimée s'appelle le facteur rural. Le facteur urbain est défini de la même façon. Ces deux facteurs sont calculés lorsque les UPE sont échantillonnées et inscrites dans chaque enregistrement de l'échantillon des secteurs NAR. Si ces facteurs sont modifiés, le changement est intégré au moment du renouvellement des UPE.

8.4 Treatment of Non Response

Some nonresponse is virtually certain to occur in any large survey. Although each interviewer is assigned a completely specified set of households in the LFS, and the interviewer makes every effort to interview these households, nonresponse will still occur for a variety of reasons, such as "no one at home" and "refusal to answer questions". The interviewer records the reason for each nonresponding household and the results are monitored on an ongoing basis. Non response is also discussed in Chapters 7 and 9.

If a nonresponding household in the current month was a respondent in the previous month, then the previous data are copied to the current month with suitable transformations being applied to certain fields. For estimation purposes, these households are treated like responding households. This imputation procedure is only used for one month; i.e., if a household is a respondent one month and a nonrespondent in the next two months, then the data are not used in the second nonresponse month.

If the above procedure cannot be used to handle nonresponse, then a weight adjustment, called a **balancing factor** is used. Balancing is carried out within each **balancing unit**. In SR areas, each subunit is a balancing unit. In NSR areas, each sampled PSU is divided into an urban and a rural balancing unit. For each balancing unit, the balancing factor is calculated as the number of households that should have been interviewed divided by the number actually interviewed (including those that received last month's data). This factor is entered on each sampled record in the balancing unit.

The use of balancing units in estimation assumes that the households that have been interviewed represent the non responding households; i.e., respondents and nonrespondents are assumed to have similar characteristics of interest, such as unemployment rate. If this assumption is not valid, then the estimates will be biased, and the bias will increase with a higher rate of nonresponse. Because the exact magnitude of this bias is impossible to calculate, every effort is made to reduce nonresponse in the field.

An unusually high nonresponse rate in a balancing unit can result in an unacceptably high balancing factor. To avoid this, the balancing unit is collapsed with a neighbouring unit whenever the balancing factor exceeds 3, and a new balancing factor is then calculated for the combined units.

8.4 Traitement de la non-réponse

Bien que, dans l'EPA, on attribue à chaque intervieweur un nombre précis de ménages à interviewer et que les intervieweurs fassent leur possible pour interviewer tous les ménages, il y a inévitablement des cas de non-réponse. Les raisons varient : certaines personnes sont absentes de chez elles et d'autres refusent de répondre. Les intervieweurs notent la raison pour laquelle un ménage n'a pas fourni de réponses, et le suivi est effectué en conséquence. La non-réponse est également traitée aux chapitres 7 et 9.

Si un ménage ne fournit pas de réponses pendant le mois en cours alors qu'il en avait fourni pour le mois précédent, les données du mois en cours sont obtenues à partir des données historiques (mois précédent) après que les corrections nécessaires soient apportées à certaines zones. Aux fins de l'estimation, ces ménages sont traités comme des ménages répondants. On n'applique cette méthode d'imputation que pour un mois. Autrement dit, si un ménage fournit des réponses un mois mais pas les deux mois suivants, on ne reprend pas les données pour le deuxième mois de non-réponse.

Si on ne peut pas avoir recours à cette méthode pour régler le problème de la non-réponse, on modifie le poids au moyen d'un facteur qu'on appelle **facteur de compensation**. La compensation est effectuée dans chaque **unité de compensation**. Dans les secteurs AR, chaque sous-unité est une unité de compensation. Dans les secteurs NAR, chaque UPE échantillonnée est divisée en une unité de compensation rurale et une unité de compensation urbaine. Pour chaque unité de compensation, on calcule le facteur de compensation en divisant le nombre de ménages qu'il aurait fallu interviewer par le nombre de ménages effectivement interviewés (qui comprend les ménages auxquels on a imputé les données du mois précédent). Ce facteur est inscrit dans chaque enregistrement échantillonné dans l'unité de compensation.

L'utilisation d'unités de compensation pour l'estimation suppose que les ménages interviewés sont représentatifs des ménages non répondants, c'est-à-dire qu'on part du principe que les ménages répondants et les ménages non répondants sont semblables du point de vue des caractéristiques qui nous intéressent, comme le chômage. Si cette hypothèse n'est pas valable, les estimations sont biaisées et le biais augmente proportionnellement au taux de non-réponse. Comme il est impossible de calculer avec exactitude l'ampleur de ce biais, on s'efforce de réduire la non-réponse sur le terrain.

D'un taux de non-réponse exceptionnellement élevé dans l'unité de compensation peut résulter un facteur de compensation si élevé qu'il est inacceptable. Pour éviter cette situation, aussitôt que le facteur de compensation d'une unité dépasse 3, celle-ci est rattachée à une unité voisine et un nouveau facteur de compensation est calculé pour la nouvelle unité ainsi formée.

8.5 Cluster Subweight

In the LFS, clusters have been designed to yield a typical sample take of three to ten households, with the lower values in urban areas. Because of rapid population growth in a cluster, the actual take may be quite different. Substantial growth can be withstood in an isolated cluster with a small expected sample before the additional take presents a field problem. However, if growth takes place in more than one cluster in an assignment, subsampling may be needed to avoid disruptions in field operations. In this case, rather than enumerate all the households that were originally selected in a high growth cluster, the inverse sampling ratio of the cluster is modified, say to k times its original value. As a result, only one out of every k originally chosen households is selected. The records for these households are then weighted by an additional factor k , the cluster subweight, since each of these households represents k times as many households as was expected by design.

In addition to their use in the subsampling of growth areas, cluster subweights can also be used in the SRU Update Program. In this program, cluster inverse sampling ratios are recalculated following an update to the cluster counts in a subunit. New inverse sampling ratios are available immediately, but it can take six to ten months before they are introduced into the active sample (because of the lead time required to introduce clusters). Thus, an interim weighting procedure, applied at the subunit level, modifies the cluster subweights to adjust for the new expected sample from the subunit. In the SRU update procedure, a change in weight at the subunit level is accomplished by modifying the cluster subweights. In general, any weighting adjustments that cannot be applied at other levels will be applied as cluster subweights. Thus, the use of cluster subweights make the LFS more flexible.

8.6 Stabilization Factor

Within each province, stabilization areas, consisting of groups of strata, have been formed. For each stabilization area, the stabilization factor is calculated as the ratio of the actual number of households in the sample to the expected number of households in the sample. If this number is less than one, then the stabilization factor is set to one. For further details, including the rationale for this factor, see Appendix B.

8.7 Estimation Methodology

The subweight is the product of the five factors discussed so far. In the final step, auxiliary information is used to transform the subweight into the final weight. Since January 1989, this has been done using a linear regression model, i.e., a model relating a dependent or response variable to several independent or explanatory variables. For the Labour Force Survey, the dependent variable can be any variable of interest and

8.5 Sous-poids de grappe

Dans l'EPA, les grappes ont été conçues de manière à ce qu'elles aient en général un rendement de l'échantillon de trois à dix ménages, les régions urbaines ayant le plus petit rendement. Comme la population peut s'accroître rapidement dans une grappe, le rendement réel peut être très différent du rendement prévu. On peut tolérer une assez forte croissance avant que celle-ci ne pose de problème sur le terrain s'il s'agit d'une grappe isolée dont le rendement prévu au départ était petit. Toutefois, si la population de plusieurs grappes d'une même tâche augmente, il se peut qu'on sous-échantillonne ces dernières pour que les opérations sur le terrain ne soient pas perturbées. A ce moment-là, au lieu d'interviewer tous les ménages choisis initialement dans chacune des grappes à croissance élevée, on modifie la fraction de sondage inverse, la ramenant à disons k fois sa valeur d'origine. Par conséquent, seulement un des k ménages choisis initialement est échantillonné. On introduit dans l'enregistrement de ces ménages un nouveau facteur de pondération k , le sous-poids de la grappe, puisque chacun de ces ménages représente k fois plus de ménages que le nombre prévu dans le plan de sondage.

Les sous-poids de grappe sont aussi utilisés dans le cadre du programme de mise à jour des UAR. Dans ce programme, on recalcule les fractions de sondage inverses de grappe après la mise à jour du chiffre de population des grappes dans une sous-unité. Les nouvelles fractions de sondage inverses sont établies immédiatement, mais il faut compter de six à dix mois avant qu'elles ne soient introduites dans l'échantillon actif (dû au délai d'introduction des grappes). On procède donc à une pondération provisoire, au niveau de la sous-unité, pour modifier les sous-poids de grappe en fonction de la taille prévue du nouvel échantillon qui sera prélevé dans la sous-unité. Selon la méthode de mise à jour des UAR, on change les poids au niveau des sous-unités en modifiant les sous-poids de grappe. En général, lorsqu'on ne peut pas modifier les poids à un autre niveau on le fait au niveau des sous-poids de grappe. L'utilisation des sous-poids de grappe ajoute donc à la souplesse du plan de l'EPA.

8.6 Facteur de stabilisation

On a formé dans chaque province des secteurs de stabilisation, qui sont constitués de groupes de strates. Le facteur de stabilisation de chaque secteur de stabilisation est le ratio du nombre réel au nombre prévu de ménages dans l'échantillon. Si ce nombre est inférieur à 1, le facteur de stabilisation est fixé à 1. Pour plus de détails au sujet de ce facteur, notamment la raison pour laquelle on y a recours, voir l'annexe B.

8.7 Méthode d'estimation

Le sous-poids est le produit des cinq facteurs dont nous avons parlé jusqu'à présent. A la dernière étape, on a recours à des informations supplémentaires pour transformer le sous-poids en un poids final. Depuis janvier 1989, on le fait en utilisant un modèle de régression linéaire, c'est-à-dire un modèle qui met en rapport une variable dépendante ou variable de réponse et plusieurs variables indépendantes ou variables explicatives. Pour l'enquête sur la population active,

the independent variables are the auxiliary variables: age-sex groups, economic regions, census metropolitan areas. The regression is set up to ensure that the final weights it produces sum to the census projections for the auxiliary variables. For example, the sum of all the final weights for females in the sample who are between 30 and 34 years will equal the census projection for that age-sex group. It is important to note that regression is not used to produce estimates directly – it is merely a convenient and flexible tool for producing appropriate final weights, which are defined implicitly in the regression model. The previous method used in the LFS for using auxiliary information is discussed at the end of this section.

Once the final weights are obtained, an estimate for an item can be obtained by summing (*final weight*) \times (*item*) for all appropriate sample members. For example, if we want to estimate the number of women aged 65 and over who are employed, then we set item to 1 for individuals who are employed and to 0 otherwise. We then take the above sum over all women in the sample who are at least 65 years old.

Estimates of ratios, such as the unemployment rate, are calculated as ratios of two estimates. For example, the provincial unemployment rate for males is estimated by the ratio of the estimated number of unemployed males to the estimated number of males in the labour force. The numerator and denominator are calculated as in the previous paragraph.

Previous LFS Estimation Methodology

During the period from January 1985 to December 1988, Labour Force Survey estimates were obtained using a raking ratio estimation procedure, which is an alternative to the weight adjustment method described above. In this procedure, the auxiliary variables used for adjustment were the Census projections for the population aged 15 years and older in (1) the CMAs, (2) the non-CMA portions of economic regions, and (3) 24 age-sex groups.

A two iteration procedure was used, with each iteration consisting of two steps.

First, the subweights were adjusted to sum to the CMA and non-CMA Economic Region counts. Then, the revised weights were adjusted to meet the age-sex group projections. These two steps were repeated once more to obtain the final weights used in estimation and tabulation.

Variances for the raking ratio estimates were obtained using a linearization method. This is described by G. H. Choudhry and H. Lee (1987).

la variable dépendante peut être n'importe laquelle des variables qui nous intéressent et les variables indépendantes sont les variables auxiliaires : groupes d'âge-sexe, régions économiques et régions métropolitaines de recensement. Le but de la régression est de faire en sorte que la somme des poids finals produits soit égale aux projections du recensement pour les variables auxiliaires. Par exemple, la somme des poids finals pour les femmes dans l'échantillon ayant entre 30 et 34 ans sera égale aux projections du recensement pour ce groupe d'âge-sexe. Il importe de noter qu'on ne se sert pas de la régression pour produire des estimations directement; c'est plutôt un outil pratique et souple qui permet de produire les bons poids finals, lesquels sont définis implicitement dans le modèle de régression. La méthode utilisée antérieurement pour l'EPA et qui avait recours elle aussi à des informations supplémentaires est présentée à la fin de cette section.

Une fois qu'on a obtenu les poids finals, on peut produire une estimation relative à une variable en faisant la somme du produit (*poids final*) \times (*variable*) pour tous les membres de l'échantillon visés. Ainsi, si l'on veut estimer le nombre de femmes âgées de 65 ans et plus qui sont occupées, on pose la variable égale à 1 pour les personnes occupées et à 0 pour celles qui ne le sont pas. On fait ensuite la somme décrite plus haut pour toutes les femmes dans l'échantillon qui ont au moins 65 ans.

Pour calculer les estimations de ratios comme le taux de chômage, on fait le ratio de deux estimations. Par exemple, pour estimer le taux de chômage provincial chez les hommes, on fait le ratio du nombre estimé d'hommes en chômage au nombre estimé d'hommes dans la population active. Le numérateur et le dénominateur sont calculés comme dans le paragraphe précédent.

Ancienne méthode d'estimation de l'EPA

De janvier 1985 à décembre 1988, les estimations de l'enquête sur la population active ont été produites au moyen d'une technique d'estimation itérative par le quotient, plutôt qu'au moyen de la méthode de rajustement des poids décrite ci-dessus. Selon cette technique, les variables auxiliaires utilisées pour rajuster les poids étaient les projections du recensement relatives à la population âgée de 15 ans et plus des RMR, des parties des régions économiques autres que les RMR et des 24 groupes d'âge-sexe.

La technique d'estimation était une technique à deux itérations, chaque itération se faisant en deux étapes.

Premièrement, les sous-poids étaient corrigés en fonction de la somme des chiffres de population pour les RMR et les parties des RE autres que les RMR. Ensuite, les poids révisés étaient corrigés pour correspondre aux projections concernant les groupes d'âge-sexe. Ces deux étapes étaient répétées pour obtenir les poids finals utilisés dans les estimations et les totalisations.

Pour obtenir la variance des estimations itératives par le quotient, on a eu recours à une méthode de linearisation. Cette méthode est décrite par G.H. Choudhry et H. Lee (1987).

8.8 Algebraic Description of Weighting

The basic weight for records in stratum h , in economic region e , in province p will be denoted by W_{peh} . This weight will be modified by the following factors to produce the subweight.

Let F_{pej} denote the rural-urban factor for province p , economic region e and type of area j . For self-representing areas ($j = 1$), $F_{pej} = 1$. We will now define F for non- self-representing areas.

Let $P_{(81)pehi}$ denote the 1981 census count of persons in province p , economic region e , stratum h , primary sampling unit i and type of area j . In non-self representing (NSR) areas, the probability π_{pehi} of selecting primary sampling unit (PSU) is proportional to its 1981 census population count and is equal to

$$\pi_{pehi} = n_{peh} \frac{P_{(81)pehi}}{P_{(81)pe..}}$$

where n_{peh} is the number of PSUs selected in peh , and a subscript '.' denotes summation over an index. Then the estimated 1981 population in province p , economic region e in type of area j , based on the sampled PSUs, is

$$\hat{P}_{(81)pe..j} = \sum_{he} \sum_i \frac{P_{(81)pehij}}{\pi_{pehi}}$$

Thus, the rural-urban factor is

$$F_{pej} = \frac{P_{(81)pe..j}}{\hat{P}_{(81)pe..j}}$$

where $j = 2$ for urban NSR areas and $j = 3$ for rural NSR areas.

Normally, the cluster subweight C is equal to 1 except where subsampling due to growth has taken place. However, the cluster subweight is sometimes used to account for other changes. See Section 8.5 for details.

The sample stabilization weight S will equal 1 unless the actual number of dwellings in a stabilization area exceeds the expected, or base, number of dwellings in the area. In the latter case, S is set equal to the ratio of actual to expected dwellings. There is one exception: if the stabilization area contains growth clusters with cluster subweights different from 1, then the sample stabilization weight for such clusters is set equal to 1, and the sample stabilization weight for the remaining clusters in the area is calculated after reducing the expected and actual number of dwellings in the area by the corresponding figures for the growth clusters.

Finally, the non response balancing factor B is calculated as

$$B_{p,e,h,i,j} = \frac{\text{expected no. of sample households in } (p,e,h,i,j)}{\text{actual no. of sample households in } (p,e,h,i,j)}$$

8.8 Description algébrique de la pondération

Soit W_{peh} le poids de base pour les enregistrements de la strate h dans la région économique e dans la province p . Pour obtenir le sous-poids, on modifiera ce poids par les facteurs qui suivent.

Soit F_{pej} le facteur rural-urbain pour la province p , la région économique e et le genre de secteur j . Pour les secteurs autoreprésentatifs, ($j = 1$), $F_{pej} = 1$. Nous allons définir à présent F pour les secteurs non autoreprésentatifs.

Soit $P_{(81)pehi}$ le nombre de personnes, selon le recensement de 1981, dans la province p , la région économique e , la strate h , l'unité primaire d'échantillonnage i et le genre de secteur j . Dans les secteurs non autoreprésentatifs (NAR), la probabilité π_{pehi} de sélection de l'unité primaire d'échantillonnage (UPE) i dans la strate h , la région économique e , la province p est proportionnelle au chiffre de population de l'UPE selon le recensement de 1981 et est égale à :

$$\pi_{pehi} = n_{peh} \frac{P_{(81)pehi}}{P_{(81)pe..}}$$

où n_{peh} est le nombre d'UPE choisies dans peh et l'indice '.' désigne une sommation sur un indice. Par conséquent, la population estimée en 1981 de la province p , la région économique e , le genre de secteur j , d'après les UPE échantillonnées est :

$$\hat{P}_{(81)pe..j} = \sum_{he} \sum_i \frac{P_{(81)pehij}}{\pi_{pehi}}$$

Le facteur rural-urbain est donc

$$F_{pej} = \frac{P_{(81)pe..j}}{\hat{P}_{(81)pe..j}}$$

où $j = 2$ pour les parties urbaines des secteurs NAR et $j = 3$ pour les parties rurales des secteurs NAR.

Normalement, le sous-poids de grappe C est égal à 1 sauf si la grappe a été sous-échantillonnée parce qu'une trop grande croissance de la population s'y est produite. Il arrive cependant qu'on se serve du sous-poids de grappe pour rendre compte d'autres changements. Voir la section 8.5 pour plus de détails.

Le poids de stabilisation de l'échantillon S est égal à 1 à moins que le nombre réel de logements dans un secteur de stabilisation ne dépasse le nombre prévu de logements dans ce secteur (ou allocation de base); à ce moment-là, on pose S égal au ratio du nombre réel au nombre prévu de logements. Il y a une exception : si le secteur de stabilisation contient des grappes où la population s'est accrue et dont le sous-poids de grappe n'est pas 1, on pose à 1 le poids de stabilisation de l'échantillon de ces grappes et on calcule le poids de stabilisation de l'échantillon pour les grappes restantes après avoir retranché du nombre prévu et du nombre réel de logements dans le secteur les nombres correspondants pour les grappes où la population s'est accrue.

Enfin, on calcule le facteur B de compensation pour la non-réponse par la formule :

$$B_{p,e,h,i,j} = \frac{n^{\circ} \text{ prévu de ménages dans } (p,e,h,i,j)}{n^{\circ} \text{ réel de ménages dans } (p,e,h,i,j)}$$

The expected number of households refers to the number of dwellings selected for a given month, minus the number of ineligible dwellings such as vacant dwellings, dwellings under construction, and so on. The actual number of households is the number of households for which a response is available either as a result of a successful interview or through the imputation procedure discussed in Section 8.4.

The subweight is the product of the basic weight and the four factors that have just been defined:

$$W^{(o)}_{pehijc} = F_{pej} C_{pehijc} S_{pehijc} B_{pehij} W_{peh}.$$

Here, c denotes cluster, and p is province, e is economic region, h is stratum, i is PSU in NSR areas and is redundant for SR areas, and j is type of area. We can now describe the transformation of the subweight into a final weight for estimation. Let y denote a variable of interest, such as whether a person is unemployed; y is an N by 1 vector with i th component equal to 1 if the corresponding individual in the population is unemployed and equal to 0 otherwise. N is the population size. Suppose that there are v auxiliary variables such as age-sex groups, economic regions and metropolitan areas available. For each individual in the population, a row vector of these v variables can in principle be created: it will consist of zeros everywhere except for the characteristics which apply to that individual. These N row vectors can be assembled into an N by v matrix Z . To take a greatly simplified example, suppose that the auxiliary variables are male, female, eastern-Canadian and western-Canadian. Then v equals 4 and Z is an N by 4 matrix. Two typical rows of Z would be (1,0,0,1) and (1,0,1,0). The first denotes a male living in western Canada and the second a male living in eastern Canada. According to the linear regression model, y and Z are related by

$$Y = Z\beta + \epsilon$$

where ϵ is a vector of error terms (deviations) and β is a vector of regression coefficients. Note that the model is for the whole population, which is unavailable. Since a sample is available, it can be used to estimate β by

$$b = (Z'WTZ)^{-1} Z'WTy,$$

where T is a diagonal matrix with i th diagonal element equal to 1 if the corresponding individual is in the sample and 0 otherwise, and W is the diagonal matrix of subweights $W^{(o)}$ (the subweight for non-sample individuals can be set to 0).

Note: This approach will lead to weights that are not the same for all household members. To ensure that there is a unique household weight, the rows of Z are modified as follows: each element of a row is replaced by the household average; for example, if the household has five members and two are females in the same age group, then every member is assigned a value of 2.5 for that auxiliary variable (age-sex group). Thus, all members of a household will have identical rows in the matrix Z .

Le nombre prévu de ménages est le nombre de logements choisis un certain mois, moins le nombre de logements qui ne sont pas admissibles comme les logements vacants, les logements en construction et ainsi de suite. Le nombre réel de ménages est le nombre de logements pour lesquels on a obtenu des réponses soit à la suite d'une interview soit par la méthode d'imputation décrite à la section 8.4.

Le sous-poids est le produit du poids de base et des quatre facteurs que nous venons de définir :

$$W^{(o)}_{pehijc} = F_{pej} C_{pehijc} S_{pehijc} B_{pehij} W_{peh}.$$

Ici, c désigne la grappe et p , la province, e , la région économique, h , la strate, i , l'UPE dans les secteurs NAR (redondant pour les secteurs autoreprésentatifs) et j , le genre de secteur. Nous pouvons à présent décrire comment le sous-poids devient le poids final aux fins de l'estimation. Soit y une variable qui nous intéresse, par exemple le fait qu'une personne soit en chômage; y est un vecteur de dimension N par 1 dont la i ème composante est égale à 1 si la personne correspondante dans la population est en chômage et égale à 0 autrement. N est la taille de la population. Supposons qu'il y a v variables auxiliaires comme les groupes d'âge-sexe, les régions économiques et les régions métropolitaines. Pour chaque personne dans la population, on peut en principe créer un vecteur ligne de ces v variables : il sera constitué de zéros partout sauf pour les caractéristiques qui s'appliquent à cette personne. On peut faire une matrice Z de N par v au moyen de ces N vecteurs lignes. Pour prendre un exemple très simplifié, supposons que les variables auxiliaires sont homme, femme, Canadien de l'Est et Canadien de l'Ouest. Alors, v est égal à 4 et Z est une matrice de N par 4. Deux rangées typiques de Z seraient (1,0,0,1) et (1,0,1,0). La première désigne un homme qui vit dans l'ouest du Canada et la seconde, un homme qui vit dans l'est du Canada. Selon le modèle de régression linéaire, le lien entre y et Z est le suivant :

$$Y = Z\beta + \epsilon$$

où ϵ est un vecteur de termes d'écarts (l'erreur) et β , un vecteur de coefficients de régression. Noter que le modèle est pour la population entière, sur laquelle on ne dispose pas de chiffres. Comme on a un échantillon, on peut s'en servir pour estimer β par la formule :

$$b = (Z'WTZ)^{-1} Z'WTy,$$

où T est la matrice diagonale dont l' i ème élément diagonal est égal à 1 si la personne en cause est dans l'échantillon et égal à 0 autrement et où W est la matrice diagonale des sous-poids $W^{(o)}$ (on peut poser à 0 le sous-poids des personnes non échantillonnées).

Nota : Cette méthode donne des poids qui ne sont pas les mêmes pour tous les membres d'un ménage. Pour faire en sorte qu'on ait un seul poids par ménage, on modifie les lignes de Z de la manière suivante : chaque élément d'une ligne est remplacé par la moyenne pour le ménage; par exemple, si le ménage est constitué de cinq membres dont deux sont des femmes du même groupe d'âge, on attribue la valeur 2.5 à chaque membre pour cette variable auxiliaire (groupe d'âge-sexe). Ainsi, les lignes dans la matrice Z seront identiques pour tous les membres du ménage.

For each auxiliary variable, a census projection of the population total for the variable is assumed to be available. Let p denote the v by 1 vector of these projected totals. Let Y denote the total for the variable of interest, that is, $Y = y'1$, where 1 is a vector of ones. The total Y can be estimated by

$$\hat{Y} = b'p = y'WTZ(Z'WTZ)^{-1}p.$$

This estimator defines implicitly the vector of final weights

$$W^{(1)} = WTZ(Z'WTZ)^{-1}p$$

since the estimator of Y can be written as $\hat{Y} = y'W^{(1)}$. Since the set of final weights $W^{(1)}$ does not depend on the characteristic Y , it can be calculated once and then used for any characteristic of interest. The paper by Lemaitre and Dufour (1987) presents a more detailed description of the material in this section.

8.9 An Example of Weighting a Record

The weighting system is an integral part of the head office processing system. The inputs to the weighting system consist of an edited unweighted file, and a file containing cluster sub-weights, rural-urban factors and census projections by age-sex and subprovincial categories. Provincial inverse sampling ratios are contained in tabular form within the actual weighting program itself. The unweighted file is first processed to calculate balancing factors for non-response. At this stage, a file of balancing factors is created and a printout of the file is generated. Balancing units with unusually high non-response are collapsed and the balancing factor file is updated according to the revised balancing factors.

For illustration, the weighting of a hypothetical record will be shown. Consider a record corresponding to a male, age 26 who was a respondent in PSU-GRCLUS 81007-21-006 in the survey month in question. Other information necessary to weight this record will be assumed at the step at which it is required. Assumed values are not necessarily the actual values of the survey.

Calculation of the balancing factor. The record under consideration is in the NSR rural portion of stratum 81000 (the first stratum in Economic Regional 1 of Alberta). The rural portions of each PSU selected in the stratum therefore form balancing units, i.e., in this case the rural portion of PSU 81007. To obtain the balancing factor it is necessary to divide the expected households (the number of households selected, excluding vacant dwellings) by the number actually interviewed or imputed on the basis of last month's records. Assume also that cluster sub-sampling takes place in cluster 81007-21-006, and that in this case the cluster inverse sampling ratio (isr) was changed from

On suppose qu'on a un chiffre de population basé sur une projection de recensement pour chaque variable auxiliaire. Soit p un vecteur de dimension v par 1 de ces totaux projetés. Soit Y le total pour la variable qui nous intéresse, c'est-à-dire $Y = y'1$, où 1 est un vecteur de uns. On peut estimer Y total par la formule :

$$\hat{Y} = b'p = y'WTZ(Z'WTZ)^{-1}p.$$

Cet estimateur définit implicitement le vecteur des poids finals

$$W^{(1)} = WTZ(Z'WTZ)^{-1}p$$

puisque l'estimateur de Y peut être formulé par l'équation $\hat{Y} = y'W^{(1)}$. Comme la série de poids finals $W^{(1)}$ ne dépend pas de la caractéristique Y , on peut la calculer une fois et s'en servir pour n'importe laquelle des caractéristiques qui nous intéressent. L'article de Lemaitre et Dufour (1987) contient une description plus détaillée de la matière présentée ici.

8.9 Comment pondérer un enregistrement : un exemple

Le système de pondération fait partie intégrante du système de traitement des données au bureau central. On introduit dans ce système un fichier de données non pondérées et vérifiées et un fichier contenant les sous-poids de grappe, le facteur rural-urbain, les projections du recensement par âge-sexe et les catégories infraprovinciales. Les fractions de sondage inverses pour les provinces sont incluses sous forme de tableaux dans le programme de pondération lui-même. La première opération à laquelle le fichier non pondéré est soumis est le calcul des facteurs de compensation pour la non-réponse. Un fichier de facteurs de compensation est créé et imprimé sur papier. Les unités de compensation où le taux de non-réponse est exceptionnellement élevé sont groupées et le fichier de facteurs de compensation est mis à jour en conséquence.

A titre d'illustration, voyons comment se fait la pondération d'un enregistrement hypothétique. Prenons le cas d'un enregistrement se rapportant à un homme de 26 ans qui aurait participé à l'enquête dans l'UPE-GROUPE-GRAPPE 81007-21-006 le mois en question. Les autres renseignements nécessaires à la pondération de cet enregistrement seront fournis à l'étape où l'on en aura besoin. Les valeurs hypothétiques ne sont pas nécessairement les valeurs réelles de l'enquête.

Calcul du facteur de compensation – L'enregistrement que nous allons étudier est dans la partie rurale d'un secteur NAR de la strate 81000 (la première strate dans la région économique (RE) 1 de l'Alberta). Les parties rurales de chaque UPE choisie dans la strate forment donc des unités de compensation, en l'occurrence la partie rurale de l'UPE 81007. Pour obtenir le facteur de compensation, il faut diviser le nombre prévu de ménages (le nombre de ménages échantillonnés moins les logements vacants) par le nombre de ménages réellement interviewés ou pour lesquels les données ont été imputées d'après les réponses fournies le mois précédent. Supposons également que la grappe 81007-21-006 est sous-échantillonnée et que la fraction de sondage inverse

one to seven. The cluster sub-weight then is seven. Further assume that in the cluster, the expected take is eight households (if the isr had not been modified, the expected take would have been approximately $7 \times 8 = 56$ households).

Of the eight households, assume six were respondents, one was not at home this month but had responded last month, and the last household was non-respondent for the second consecutive month. The response status of the first month non-respondent household for which there was a previous month's response would have been changed to a response prior to weighting, so that for the purposes of weighting the respondent households number seven, with one nonrespondent household.

In the remainder of the balancing unit (rural portion of PSU 81007), assume that the expected households were 44, the number of enumerated households equals 35, and further that the only cluster in the balancing unit with a cluster sub-weight is cluster 81007-71-006. The expected number of households according to the design in the balancing unit is then given by $44 + (\text{cluster sub-weight} \times \text{expected take}) = 44 + (7 \times 8) = 100$, and the expected number of enumerated households is given by $35 + (\text{cluster sub-weight} \times \text{enumerated households}) = 35 + (7 \times 7) = 84$ (since households in sub-sampled clusters are weighted to reflect expected sample according to the design). The balancing unit weight $B_{phij} = 100/84 = 1.19048$. If B_{phij} had been greater than three, then collapsing of units would take place and the rural portion for the whole stratum would be considered as a balancing unit for non-response, instead of each PSU being considered separately. The balancing factors would be recalculated on this basis.

Calculation of the subweight. The PSU-GR identifies the record as NSR rural for Alberta which establishes the following weights: the basic weight $W_{peh} = 125$, and the rural-urban factor for the province $F_{pej} = 0.997114$ (reflecting the fact that in the selected PSUs in Alberta at this time the rural population was slightly overrepresented). The balancing factor $B_{pehij} = 1.19048$ was calculated previously, and a cluster sub-weight of 7.0 is input and applied to the records in this cluster. Finally, assume that no stabilization takes place in the area containing this PSU, i.e., $S_{pehijc} = 1$. Thus, the subweight for the cluster is given by:

$$\begin{aligned} W_{pehij} &= W_{peh} F_{pej} C_{pehijc} S_{pehijc} \\ &= 125 \times 0.997114 \times 1.19048 \times 7 \times 1 \\ &= 148.38053 \times 7 \\ &= 1038.6637 \end{aligned}$$

Calculation of the final weight. The procedure we have just described is applied to each record on the file. The file is processed as described in Section 8.8 to produce a final weight for each record.

(fsi) est passée de un à sept. Le sous-poids de la grappe est alors de sept. Supposons enfin que le rendement prévu de l'échantillon de la grappe est de huit ménages (si la fsi n'avait pas été modifiée, le rendement prévu aurait été d'environ $7 \times 8 = 56$ ménages).

Supposons que, sur les huit ménages, six ont fourni des réponses, un n'a pas été rejoint ce mois-ci mais avait fourni des réponses le mois précédent et le dernier n'a pas fourni de réponses pour le deuxième mois consécutif. Le statut du ménage qui n'a pas fourni de réponses ce mois-ci mais qui l'avait fait le mois précédent est changé avant la pondération de sorte qu'il est considéré comme un ménage répondant; pour les besoins de la pondération, on compte donc qu'il y a sept ménages répondants et un ménage non répondant.

Supposons que, dans le reste de l'unité de compensation (partie rurale de l'UPE 81007), le nombre prévu de ménages était de 44 et le nombre de ménages interviewés, de 35 et par ailleurs que la seule grappe sous-pondérée est la grappe 81007-71-006. Le nombre prévu de ménages d'après le plan de sondage dans l'unité de compensation se calcule donc ainsi : $44 + (\text{sous-poids de la grappe} \times \text{rendement prévu}) = 44 + (7 \times 8) = 100$, et le nombre prévu de ménages interviewés se calcule ainsi : $35 + (\text{sous-poids de la grappe} \times \text{nombre de ménages interviewés}) = 35 + (7 \times 7) = 84$ (étant donné que le nombre de ménages dans les grappes sous-échantillonnées est pondéré en fonction de l'échantillon prévu selon le plan de sondage). Le poids de l'unité de compensation B_{phij} est égal à : $100/84 = 1.19048$. Si B_{phij} avait été supérieur à 3, on aurait groupé des unités, considéré la partie rurale de la strate entière comme une unité de compensation pour la non-réponse au lieu de considérer chaque UPE séparément et recalculé les facteurs de compensation sur cette base.

Calcul du sous-poids – Le code d'UPE-GR nous dit que l'enregistrement est celui d'une personne qui habite dans une UNAR rurale en Alberta, de sorte que les poids sont les suivants : le poids de base $W_{peh} = 125$ et le facteur rural-urbain pour la province $F_{pej} = 0.997114$ (ce qui montre que, à ce moment là, dans les UPE échantillonnées en Alberta, la population rurale est légèrement surreprésentée). Le facteur de compensation $B_{pehij} = 1.19048$ a déjà été calculé, et un sous-poids de grappe de 7.0 est attribué aux enregistrements dans cette grappe. Supposons enfin que le secteur dans lequel se trouve cette UPE n'a pas été stabilisé, autrement dit que $S_{pehijc} = 1$. On calcule donc le sous-poids de la grappe par la formule :

$$\begin{aligned} W_{pehij} &= W_{peh} F_{pej} C_{pehijc} S_{pehijc} \\ &= 125 \times 0.997114 \times 1.19048 \times 7 \times 1 \\ &= 148.38053 \times 7 \\ &= 1038.6637 \end{aligned}$$

Calcul du poids final – L'opération que nous venons de décrire est effectuée pour chacun des enregistrements dans le fichier. Pour obtenir le poids final pour chaque enregistrement, on traite le fichier de la manière indiquée à la section 8.8.

8.10 Variance Estimation

LFS estimates are based on a sample. If several independent samples using the LFS design were to be selected for a given month, different estimates would be obtained since each sample would be different. A measure of the variability of these estimates would provide an indication of the reliability of an estimate produced using the actual LFS sample. Because the LFS is a probability survey, it is possible to obtain such a measure, namely the **variance**, for each estimate.

Although the sampling and estimation procedures for the LFS are complex, the variance estimation method, using a technique known as the jackknife, is conceptually simple. Each cluster in SR areas and each PSU in NSR areas will be referred to as a **replicate**. Suppose that Y is the characteristic of interest and \hat{Y} is its estimate, as described in Section 8.8. Remove a replicate i from its stratum h and adjust the remaining replicates in the stratum to compensate for the removal: if there were originally, say, three replicates, then the adjustment is 3/2. Leaving all other strata with their full samples, apply the regression procedure described in Section 8.8 to obtain an estimate $\hat{Y}_{(hi)}$ of Y ; the subscript (hi) indicates that the estimate is obtained after replicate i in stratum h is removed. Now repeat these steps for every replicate in every stratum. This yields as many estimates $\hat{Y}_{(hi)}$ as there are replicates in the sample, in addition to the original LFS estimate \hat{Y} . The jackknife variance estimate of \hat{Y} is given by

$$\text{var}(\hat{Y}) = \sum_h \frac{(n_h - 1)}{n_h} \sum_i (\hat{Y}_{(hi)} - \hat{Y})^2,$$

where n_h is the number of replicates in stratum h . Like previous variance estimation procedures used in the LFS, the current estimator's validity is based on the assumption that replicates are selected independently, which is approximately true for the LFS design. The jackknife in this case normally leads to a slight overestimation of variances and, thus, is conservative.

Variance estimates are produced monthly for many characteristics at various levels, such as province, economic region and census metropolitan area. The presentation and use of these figures is discussed further in the next chapter.

8.10 Estimation de la variance

Les estimations de l'EPA sont basées sur des données recueillies auprès d'un échantillon. Si, un certain mois, on prélevait plusieurs échantillons indépendants en se servant du plan de sondage de l'EPA et qu'on produisait des estimations d'après les données recueillies auprès de chacun, ces estimations seraient différentes puisque les échantillons seraient différents. Une mesure de la variabilité de ces estimations nous donnerait une indication de la fiabilité des estimations produites à partir de l'échantillon réel de l'EPA. Comme l'EPA est une enquête à échantillonnage probabiliste, il est possible d'obtenir une telle mesure, à savoir la **variance**, pour chaque estimation.

Bien que les méthodes d'échantillonnage et d'estimation de l'EPA soient complexes, la méthode d'estimation de la variance, qui utilise une technique dite du jackknife, est simple du point de vue théorique. Chaque grappe des secteurs AR et chaque UPE des secteurs NAR est considérée comme une **réplique**. Supposons que Y est la caractéristique qui nous intéresse et que \hat{Y} est l'estimation qui s'y rapporte, comme nous l'avons vu à la section 8.8. On retire une réplique i de sa strate h et on rajuste les répliques qui restent dans la strate pour compenser; s'il y avait au départ disons trois répliques, le rajustement est de l'ordre de 3/2. En ne touchant pas aux échantillons des autres strates, on applique la méthode fondée sur la régression décrite à la section 8.8 pour obtenir une estimation $\hat{Y}_{(hi)}$ de Y ; l'indice (hi) indique qu'on a obtenu l'estimation après avoir retiré la réplique i de la strate h . Répétons cette opération pour chaque réplique dans chaque strate. On obtient autant d'estimations $\hat{Y}_{(hi)}$ qu'il y a de répliques dans la strate, en plus de l'estimation originale de l'EPA \hat{Y} . On calcule l'estimation de la variance de \hat{Y} par la technique dite du jackknife par la formule :

$$\text{var}(\hat{Y}) = \sum_h \frac{(n_h - 1)}{n_h} \sum_i (\hat{Y}_{(hi)} - \hat{Y})^2,$$

où n_h est le nombre de répliques dans la strate h . La validité de l'estimateur actuel, comme celle des méthodes d'estimations de la variance utilisées antérieurement pour l'EPA, suppose que les répliques sont prélevés de façon indépendante, ce qui est à peu près vrai pour le plan de l'EPA. Lorsqu'on utilise la technique dite du jackknife dans ce cas, l'estimation de la variance est légèrement supérieure à ce qu'elle devrait être; cette technique est donc modérée.

On produit chaque mois des estimations de la variance relativement à de nombreuses caractéristiques et pour divers niveaux géographiques comme la province, la région économique et la région métropolitaine de recensement. Il sera question plus amplement de la présentation et de l'utilisation de ces données dans le prochain chapitre.

Chapter 9

Data Quality

9.1 Introduction

To appreciate the value of data obtained from a sample survey, such as the Labour Force Survey, it is necessary to understand clearly the role of **Sampling error** and **Non-sampling error** in surveys.

In sample surveys, inference is made about the entire population covered by the survey on the basis of data obtained from only a part (sample) of that population. The results are likely to be different from those that would be obtained if the entire population were enumerated under the same survey conditions. The error arising from drawing inferences about the entire population on the basis of information from a sample is termed sampling error.

Non-sampling error, as the name implies, has nothing to do with the sampling process and is present in a Census (where all units of the population are enumerated) as well as in a sample survey. This type of error is associated with the processes of the survey at all stages, from data collection and processing through to the final tabulations.

This chapter, after a brief discussion of sampling error in sections 9.2 and 9.3, describes the sources of non-sampling errors, the means used to detect, measure and assess each type and the monitoring and corrective procedures used to minimize them.

9.2 Sampling Error

The magnitude of the sampling error for survey estimates depends on many factors. The most obvious is the size of the sample. The larger the sample size, the smaller the sampling error. For a given sample size, the sampling error will depend on various features of the sample design, such as the stratification procedure employed, the allocation of the sample, the choice of sampling units and the method of selection used at the different stages of sampling in a multi-stage design. In addition, for a given sample design, different estimation procedures will result in different sampling errors. Finally, given a common sample size, sample design and estimation procedure, estimates of different characteristics will generally have different sampling errors. Characteristics that are relatively scarce or unevenly distributed in the population will tend to have larger sampling errors than characteristics

Chapitre 9

Qualité des données

9.1 Introduction

Pour apprécier la qualité des données recueillies au moyen d'une enquête par sondage comme l'enquête sur la population active, il faut bien comprendre le rôle que jouent **l'erreur d'échantillonnage** et **l'erreur non due à l'échantillonnage** dans ce genre d'enquête.

Dans une enquête par sondage, des inférences sont faites au sujet de la population visée entière à partir de données recueillies auprès d'une partie seulement (d'un échantillon) de cette population. Les résultats sont probablement différents de ceux qu'on obtiendrait si on menait un recensement complet auprès de cette population dans les mêmes conditions que celles de l'enquête. L'erreur d'échantillonnage découle du fait que les inférences portant sur la population entière sont basées uniquement sur un échantillon provenant de cette population.

L'erreur non due à l'échantillonnage, comme son nom l'indique, n'a rien à voir avec le processus d'échantillonnage et se produit dans un recensement (auquel participent toutes les unités de la population) aussi bien que dans une enquête par sondage. Ce genre d'erreur est associé à toutes les étapes de l'enquête, de la collecte au traitement des données et même jusqu'à la production des totalisations finales.

Dans ce chapitre, nous allons d'abord parler brièvement de l'erreur d'échantillonnage (sections 9.2 et 9.3), puis nous allons décrire les sources des erreurs non dues à l'échantillonnage, les moyens utilisés pour détecter, mesurer et évaluer chaque catégorie d'erreur de ce genre et les mesures en place (vérification, correction) visant à en réduire le nombre au maximum.

9.2 Erreur d'échantillonnage

L'ampleur de l'erreur d'échantillonnage dont sont entachées les estimations d'enquête est fonction d'un grand nombre de facteurs. Le plus évident est la taille de l'échantillon. Plus un échantillon est grand, plus l'erreur est petite. Pour un échantillon d'une taille donnée, l'erreur d'échantillonnage est liée à diverses caractéristiques du plan de sondage comme la méthode de stratification utilisée, la répartition de l'échantillon, le choix des unités d'échantillonnage et la méthode de sélection employée à chaque degré d'échantillonnage s'il s'agit d'un plan à plusieurs degrés. En outre, pour un plan de sondage donné, différentes méthodes d'estimation entraîneront des erreurs différentes. Enfin, même si la taille de l'échantillon, le plan de sondage et la méthode d'estimation sont pareils, les estimations relatives aux différentes caractéristiques sont généralement entachées d'erreurs d'échantillonnage différentes. Ces erreurs sont habituellement plus grandes pour les caractéristiques relativement rares ou

that are common or widely distributed. For example, the sampling error for estimates of unemployment will generally be higher than for estimates of employment, although both are estimated from the same sample.

The efficiency of a sample design and estimation procedure is usually measured by the **mean square error** of one or more characteristics. The mean square error is defined as the average of the square of deviations of the estimate of the characteristic from the "true value" taken over all possible samples. The average of the estimates over all possible samples is termed the **expected value** of the estimate. The difference between the expected value and the true value is termed the **bias** of the estimate. The **variance** of the sample estimate is the average of the squared difference of the estimate from its expected value. The square root of the variance is called the **standard deviation** (or standard error) of the estimate.

If the estimation procedure were unbiased, the expected value and the true value of the estimate would be the same and thus the mean square error and the variance would be the same. Certain estimation procedures (such as the one used in the LFS), though negligibly biased, yield smaller mean square errors than other unbiased procedures.

The most important feature of a probability sample is that the variance of the estimate (and hence the standard deviation) can be estimated from the sample itself.

An additional measure of quality derived from the sample is known as the **design effect**. It is defined as the ratio of the variance of an estimate derived from the sample survey having a particular design to the variance of the estimate derived from the same sample survey – but assuming a simple random sample design. The design effect provides a comprehensive measure of the combined effect of all the design features, such as stratification, multistage sampling and estimation. Design effects can serve as an index to measure the deterioration of the overall design over time. The lower the design effect the more efficient the design. Measures of design effects, as well as other information, are used in determining areas for design updates discussed earlier in Section 3.4. Some representative design effects are shown in the table.

réparties inégalement dans la population que pour les caractéristiques fréquentes ou réparties de façon homogène. Les estimations du chômage, par exemple, sont généralement entachées d'une plus grande erreur d'échantillonnage que les estimations de l'emploi, bien que les deux soient basées sur le même échantillon.

On se sert habituellement de l'**erreur quadratique moyenne** d'une ou de plusieurs caractéristiques pour mesurer l'efficacité du plan de sondage et de la méthode d'estimation. L'erreur quadratique moyenne est définie comme étant la moyenne des carrés des écarts de la valeur estimée de la caractéristique par rapport à sa valeur réelle pour tous les échantillons possibles. La moyenne des estimations pour tous les échantillons possibles s'appelle l'**espérance mathématique** de l'estimation. La différence entre l'espérance mathématique et la valeur réelle s'appelle le **biais** de l'estimation. La **variance** de l'estimation d'échantillon est la moyenne des carrés des écarts de l'estimation par rapport à son espérance mathématique. On appelle la racine carrée de la variance l'**écart type** (ou erreur type) de l'estimation.

Si la méthode d'estimation n'était pas biaisée, l'espérance mathématique et la valeur réelle de l'estimation seraient identiques, et par conséquent l'erreur quadratique moyenne et la variance seraient identiques également. Certaines méthodes d'estimation (comme celle dont on se sert pour l'EPA), bien que causant un biais négligeable, donnent lieu à de plus petites erreurs quadratiques moyennes que d'autres méthodes non biaisées.

La caractéristique la plus importante d'un échantillon probabiliste est que la variance de l'estimation (et donc l'écart type) peut être estimée à partir de l'échantillon lui-même.

Il y a une autre mesure de la qualité qui s'obtient à partir de l'échantillon, et c'est l'**effet du plan**. On le définit comme étant le rapport entre la variance d'une estimation tirée d'une enquête par sondage ayant un plan donné et la variance de l'estimation tirée de la même enquête pas sondage, mais dont la méthode d'échantillonnage serait la méthode aléatoire simple. L'effet du plan donne une mesure globale de l'effet combiné de toutes les caractéristiques du plan, comme la stratification, l'échantillonnage à plusieurs degrés et l'estimation. L'effet du plan peut servir d'indice de la détérioration du plan avec le temps. Par exemple, une diminution de l'effet du plan indique que le plan devient plus efficace. On a recours à la mesure de l'effet du plan ainsi qu'à d'autres renseignements pour décider des secteurs où la mise à jour du plan, dont nous avons parlé à la section 3.4, est nécessaire. Le tableau qui suit présente quelques valeurs représentatives de l'effet du plan pour la caractéristique chômage.

Design Effects, Unemployed – August 1989

Effet du plan, chômeurs – août 1989

Province	All	Male	Female
	Les deux sexe	Hommes	Femmes
Newfoundland – Terre-Neuve	1.40	1.34	1.33
Prince Edward Island – Île-du-Prince-Édouard	1.75	1.36	1.04
Nova Scotia – Nouvelle-Écosse	1.20	1.09	1.04
New Brunswick – Nouveau-Brunswick	1.87	1.41	1.18
Quebec	1.61	1.54	1.25
Ontario	1.27	1.27	1.38
Manitoba	2.09	1.74	1.61
Saskatchewan	1.24	1.19	1.29
Alberta	1.55	1.52	1.48
British Columbia – Colombie-Britannique	1.80	1.69	1.65
Canada	1.49	1.45	1.35

9.3 Presentation of Sampling Errors

The estimated standard error can be used to provide a confidence interval of an estimate in the following way. If the estimated value of a characteristic in some area is X and the standard error of that estimate is d , then, in many cases the interval $X \pm d$ will contain the "true" value about 68 times out of a hundred and the interval $X \pm 2d$ will contain the true value 95% of the time. It is often convenient to express the standard error as a percentage of the estimate and this is known as the **coefficient of variation (CV)**.

Rough confidence intervals for the LFS estimates may be obtained from the letter symbols given in many of the tables in the monthly LFS publications. The letter symbols refer to ranges of percentage coefficient of variation for the characteristics. The letter symbol gives the user an indication of the reliability of the estimate. The ranges of values for the symbols are as follows:

A (0.0 - 0.5%)	F (10.1 - 16.5%)
B (0.6 - 1.0%)	G (16.6 - 25.0%)
C (1.1 - 2.5%)	H (25.1 - 33.3%)
D (2.6 - 5.0%)	J (33.4% +)
E (5.1 - 10.0%)	

The letter values in the tables are updated twice a year and are based on a six-month average of the CVs of the characteristic. In addition to these symbols, an explanatory note on reliability and other limitations of the data is provided in the publications.

The Labour Force Survey produces many thousands of estimates of labour force characteristics. These are available at national, provincial and sub-provincial levels for monthly estimates, month-to-month change, averages of levels and year-to-year change in annual averages. Due to space limitations, it is not possible to include direct estimates of CVs for all survey estimates in regular and special publications. However, look-up tables are available giving CV approximations for various groupings of estimates.

9.3 Présentation des erreurs d'échantillonnage

On peut se servir de l'erreur type estimée pour obtenir l'intervalle de confiance d'une estimation; il s'agit de procéder de la manière suivante. Si la valeur estimée d'une caractéristique dans un certain secteur est X et que l'erreur type de cette estimation est d , alors, dans un grand nombre de cas, l'intervalle $X \pm d$ contiendra la valeur réelle environ 68 fois sur 100 et l'intervalle $X \pm 2d$, 95 fois sur 100. Il est souvent pratique d'exprimer l'erreur type en pourcentage de l'estimation; on parle alors de **coefficient de variation (CV)**.

Il est possible de se faire une idée approximative des intervalles de confiance des estimations de l'EPA à partir des symboles en caractères alphabétiques indiqués dans un bon nombre des tableaux contenus dans les publications mensuelles se rapportant à cette enquête. Les symboles en question indiquent dans quel intervalle se situe le coefficient de variation pour la caractéristique. Ils renseignent le lecteur sur la fiabilité de l'estimation. Les tranches de valeurs correspondant à chaque symbole sont les suivantes :

A (0.0 - 0.5%)	F (10.1 - 16.5%)
B (0.6 - 1.0%)	G (16.6 - 25.0%)
C (1.1 - 2.5%)	H (25.1 - 33.3 %)
D (2.6 - 5.0%)	J (33.4% +)
E (5.1 - 10.0%)	

Deux fois par an, on révisé les tableaux et on détermine s'il y a lieu de changer les lettres compte tenu de la moyenne des coefficients de variation pour les caractéristiques sur les six derniers mois. Les publications contiennent, outre ces symboles, une note explicative sur la fiabilité des données et les lacunes que celles-ci pourraient comporter.

On produit des milliers d'estimations des caractéristiques de la population active à partir des résultats de l'enquête sur la population active. Il y a des estimations mensuelles, des estimations de la variation d'un mois à l'autre, des estimations des moyennes de niveau et des estimations de la variation des moyennes annuelles, et ce au niveau national, provincial et infraprovincial. L'espace disponible dans les publications normales et spéciales ne permet pas d'inclure des estimations directes des CV pour toutes les estimations de l'enquête. Il existe cependant des tables de recherche qui contiennent des approximations des CV pour divers groupes d'estimations.

9.4 Non-Sampling Errors

Non-sampling error (like sampling error) may be divided into two components, variance and bias. Non-sampling variance has a tendency to be in either direction and thus the net effect of this component may be negligible over many observations or for large areas; but for small areas or rare or sensitive characteristics, it can be large. An example of this would be checking one response box by mistake when another was reported. The non-sampling bias, on the other hand, has a tendency to occur in one direction more than the other. For example, it may be the result of such things as the training or attitude of the interviewer, a fault in questionnaire design, or the method of imputation for non-response, that may tend to invite error in one direction more than the other.

Contributions to non-sampling variance and bias may arise through one or more of the following main sources of error:

Coverage Error

Coverage errors are introduced whenever the frame of sample units does not adequately represent the target population at the time of the survey. These errors are caused by (1) units in the target population that are missed from the sampling frame (under coverage), (2) units included in the sampling frame that are not in the target population (over coverage), and (3) units in the target population that are included more than once in the sampling frame (duplication).

Coverage errors may occur at several stages of the survey, such as during frame preparation, sampling unit definition, assigning of the selection probabilities for sampling purposes or at the data collection and processing stages. The net under coverage or over coverage in the LFS is estimated by the **Slippage** rate. Slippage is the difference between population estimates obtained by summing subweighted LFS sample records and population estimates derived from independent sources such as the Census.

Coverage errors in the LFS generally result in a net under coverage of the population. To minimize any bias resulting from this, the sample estimates are adjusted to external control totals in the estimation procedure – see chapter 8. The result is that the LFS estimates obtained by weighting up the sample counts in various age-sex groups, economic regions and CMAs are equal to corresponding population totals from the most recent Census-based population projections. Average slippage rates and estimated standard deviations of the rates for the 1988 calendar year are given in the following table.

9.4 Erreurs non dues à l'échantillonnage

L'erreur non due à l'échantillonnage a, comme l'erreur d'échantillonnage, deux composantes : la variance et le biais. La variance non due à l'échantillonnage va généralement dans un sens ou dans l'autre de sorte que son effet net peut être négligeable si le nombre d'observations est élevé ou s'il s'agit de grands secteurs; toutefois, il peut être élevé lorsqu'il s'agit de petits secteurs ou bien lorsque les caractéristiques à l'étude sont rares ou rattachées à des questions délicates. Un exemple de ce genre d'erreur se produit lorsqu'un intervieweur coche une case ne correspondant pas à la réponse donnée par le répondant. Le biais non dû à l'échantillonnage a, pour sa part, tendance à se produire dans un sens plus que dans l'autre. Il peut être attribuable à la formation ou à l'attitude de l'intervieweur, à une faille dans la conception du questionnaire ou à la méthode d'imputation utilisée pour remédier à la non-réponse. L'ensemble de ces facteurs peuvent contribuer à provoquer l'accumulation des erreurs dans une direction plus que dans l'autre.

Les catégories d'erreurs suivantes peuvent contribuer à la variance non due à l'échantillonnage et au biais.

Erreur de couverture

Il se produit des erreurs de couverture lorsque la population cible n'est pas représentée adéquatement dans la base des unités d'échantillonnage au moment de l'enquête. Il se peut que des unités aient été omises de la base de sondage (sous-dénombrement), que des unités ne se trouvant pas dans la population cible y aient été incluses (surdénombrement) ou que des unités s'y retrouvent plus d'une fois (répétition).

Les erreurs de couverture peuvent se produire à plusieurs étapes de l'enquête : pendant la constitution de la base de sondage, la définition des unités d'échantillonnage, l'attribution des probabilités de sélection aux fins de l'échantillonnage ou encore la collecte ou le traitement des données. Le taux de glissement donne une estimation du sous-dénombrement ou du surdénombrement net de l'EPA. Il mesure l'écart entre les estimations de la population que l'on obtient en additionnant les enregistrements sous-pondérés de l'échantillon de l'EPA et les estimations de la population provenant de sources indépendantes comme le recensement.

Les erreurs de couverture de l'EPA donnent habituellement lieu à un sous-dénombrement de la population. Pour réduire au maximum le biais qui en résulte, au cours du processus d'estimation, on corrige les estimations d'échantillon en fonction de totaux de contrôle provenant d'autres sources que l'EPA (voir le chapitre 8). Par conséquent, les estimations de l'EPA que l'on obtient en pondérant les données d'échantillon pour les divers groupes d'âge-sexe, régions économiques et RMR sont égales aux chiffres de population correspondants que l'on a d'après les projections de la population fondées sur le dernier recensement. Le tableau qui suit contient les taux de glissement moyens et les écarts types estimés des taux pour l'année civile 1988.

Average Slippage Rates and Standard Deviations -1988

Canada (With Age Groups) and Provinces

Taux de glissement moyens et écarts types - 1988

Canada (selon les groupes d'âge) et provinces

		Slippage	Standard Deviation
		Glissement	Ecart type
Canada	Overall - Global	4.93	0.80
	15-19 years - ans	4.54	1.78
	20-24 years - ans	9.49	1.65
	25-44 years - ans	5.53	1.20
	45-64 years - ans	3.74	1.09
	65+ years - ans et plus	1.99	1.73
Newfoundland - Terre-Neuve		7.96	1.86
Prince Edward Island - Île-du-Prince-Édouard		7.35	2.24
Nova Scotia - Nouvelle-Écosse		6.10	1.95
New Brunswick - Nouveau-Brunswick		6.00	1.55
Quebec		2.26	1.77
Ontario		5.67	1.53
Manitoba		6.09	1.83
Saskatchewan		7.94	1.62
Alberta		6.18	1.75
British Columbia - Colombie-Britannique		4.99	2.10

Non-response Error

Each month, during the designated survey week, every reasonable effort is made to interview all eligible persons in selected dwellings. For various reasons this is not always possible. These non-interviews occur for the following reasons:

- Household non-response: This includes no one at home, temporary absence, interview impossible (weather conditions, unusual circumstances in the household, etc.) and refusals.
- Vacant dwellings: This includes unoccupied dwellings, seasonally occupied dwellings, dwellings under construction and those occupied by persons not eligible to be interviewed.
- Non-existent dwellings: This includes demolished dwellings, dwellings converted to business premises, mobile homes that had been relocated or dwellings that had been abandoned or listed in error.

Data for non-responding units are imputed using data from similar responding units or using earlier data from the non-responding units, if available. The extent of any bias due to imputation is usually unknown and is very much dependent on characteristic differences between the responding and non-responding groups of units. The impact of this bias increases with increases in the non-response rate and so attempts are made to maintain a response rate as high as possible.

Vacant and non-existent dwellings do not contribute to any bias in the sample but do result in a higher sample variance due to a smaller household count. Some typical non response rates are given below.

Non-reponse

Chaque mois, pendant la semaine d'enquête choisie, on s'efforce d'interviewer toutes les personnes admissibles dans les logements échantillonnés. Il n'est pas toujours possible de le faire, et ce pour les raisons suivantes :

- Non-réponse du ménage : Personne à la maison, absence temporaire, interview impossible (mauvais temps, circonstances inhabituelles dans le ménage, etc.) ou refus.
- Logement vacant : Logement non occupé, logement saisonnier, logement en construction ou logement occupé par des personnes ne devant pas être interviewées.
- Logement non existant : Logement démoli, logement transformé en local d'affaires, maison mobile démenagée ou encore logement abandonné ou inscrit par erreur.

Pour les unités non répondantes, les données manquantes sont déterminées à partir des données provenant d'unités semblables ou des données fournies antérieurement par les unités elles-mêmes (si elles existent). L'importance du biais attribuable à l'imputation est habituellement inconnue, mais on sait qu'elle est étroitement liée aux différences de caractéristiques entre les groupes d'unités répondantes et les groupes d'unités non répondantes. Comme l'effet de ce biais s'accroît avec l'augmentation du taux de non-réponse, on essaie de maintenir le taux de réponse à un niveau aussi élevé que possible.

Les logements vacants et non existants ne contribuent pas au biais dans l'échantillon, cependant, ils produisent une hausse de la variance de l'échantillon puisqu'ils réduisent le nombre de logements. Le tableau qui suit contient des taux de non-réponse typiques pour l'EPA.

Household Non Response Rates
Canada and Provinces – 1988

Taux de non-réponse globaux
Canada et provinces – 1988

	Average	High	Low
	Moyenne	Maximum	Minimum
Canada	5.08	6.40	4.40
Newfoundland – Terre-Neuve	4.69	7.70	3.30
Prince Edward Island – Ile-du-Prince-Édouard	3.36	5.10	2.00
Nova Scotia – Nouvelle-Ecosse	5.63	7.30	4.70
New Brunswick – Nouveau-Brunswick	4.98	6.30	3.60
Quebec	4.91	7.30	4.00
Ontario	4.76	6.40	3.80
Manitoba	5.36	6.90	4.20
Saskatchewan	5.26	6.70	3.90
Alberta	5.23	6.30	4.10
British Columbia – Colombie-Britannique	5.63	6.70	4.20

Refusal rates in the LFS are generally very low. The monthly rates in 1989 were usually between 1.1 and 1.3 percent, with a maximum of 1.6 percent. Provincial refusal rates usually fall in the same range, but may be as low as 0.8 percent and as high as 2.0 percent.

Response Error

This error may be due to questionnaire design, the wording of the questions, the attitude of the respondent, the interviewing approach or other general conditions under which the survey is conducted. Response errors may occur when data are requested, provided, received or recorded. Questions may be misinterpreted by the respondent who may not know, may not remember or may purposely wish to distort the correct answer. Different interviewers may have different tendencies in explaining questions or interpreting responses, and so on. Response errors, like other errors, can have both a variance and a bias component. Estimates of **response variance** and **response bias** can be obtained from the Reinterview Program, which is discussed in Section 9.5.

Proxy response, which is the collection of information pertaining to other household members from one member of the household, can also lead to response errors.

In repeated surveys, where the sample is made up of a number of independent panels or rotation groups, the expected values of estimates differ slightly from one rotation group to the next. This is known as **rotation group bias**. In the LFS, this effect is most pronounced for the one-sixth of the sample that are first time interviews. However, the rotation scheme ensures that the impact of rotation group bias is minimized by having five-sixths of the sample in common from one month to the next. A rotation group index can be calculated as the ratio of an estimate calculated from that portion of the sample in the survey for a particular time (first month, second month etc.) to

Les taux de refus pour l'EPA sont habituellement très bas. En général, les taux mensuels en 1989 ont varié entre 1.1 et 1.3 pour cent avec un taux maximum observé de 1.6 pour cent. Les taux de refus au niveau provincial sont ordinairement du même ordre de grandeur mais ils peuvent descendre aussi bas que 0.8 pour cent ou monter aussi haut que 2.0 pour cent.

Erreur de réponse

Cette erreur peut être attribuable à la conception du questionnaire, à la formulation des questions, à l'attitude du répondant, à la façon dont l'interview est menée ainsi qu'au conditions générales dans lesquelles l'enquête est réalisée. Les erreurs de réponse peuvent se produire au moment où l'on demande les renseignements ou au moment où les renseignements sont fournis, reçus ou inscrits. Le répondant peut mal interpréter la question, ne pas savoir la réponse, l'avoir oubliée ou préférer communiquer une distorsion des faits pour des raisons qui lui sont propres. De plus, il arrive que les différents intervieweurs aient tendance à expliquer les réponses ou à les interpréter de manière différente. Les erreurs de réponse, comme les autres catégories d'erreurs, peuvent avoir une variance et un biais. Le Programme de réinterview, dont il sera question à la section 9.5, fournit des estimations de la **variance de réponse** et du **biais de réponse**.

Les **réponses par procuration**, que l'on obtient lorsqu'on recueille auprès d'un membre du ménage des renseignements se rapportant à un autre membre, peuvent aussi entraîner des erreurs de réponse.

Dans les enquêtes répétées, où l'échantillon est constitué d'un certain nombre de panels indépendants ou groupes de renouvellement, l'espérance mathématique des estimations varie légèrement d'un groupe de renouvellement à un autre. Il se produit alors ce qu'on appelle un **biais de renouvellement**. En ce qui concerne l'EPA, ce biais atteint son plus haut niveau pour le sixième de l'échantillon qui en est à sa première interview. Cependant, le mode de renouvellement choisi, selon lequel les cinq sixièmes de l'échantillon ne changent pas d'un mois à l'autre, fait en sorte que le biais de renouvellement est réduit au maximum. On peut obtenir l'indice de renouvellement en faisant le rapport entre une estimation calculée pour la partie de l'échantillon participant à

the estimate calculated from all the sample. Average index values and estimated standard errors for the period from March 1985 to February 1988 for the three main labour force characteristics at the Canada level are shown in the table.

Rotation group bias in the LFS for earlier periods was studied by Tessier (1976) and Ghangurde (1982), following a study by Bailar (1975) of rotation group bias in the American Current Population Survey.

Average rotation group index : March 1985 – February 1988

Characteristic Caractéristique	Month in Sample – Mois de participation à l'enquête					
	First Premier	Second Deuxième	Third Troisième	Fourth Quatrième	Fifth Cinquième	Sixth Sixième
Employed – Pers. occupées	0.9745 0.0040	1.0091 0.0036	1.0102 0.0037	1.0070 0.0038	1.0028 0.0038	0.9963 0.0036
Unemployed – Chômeurs	0.9839 0.0055	1.0271 0.0065	1.0195 0.0068	1.0065 0.0066	0.9963 0.0060	0.9668 0.0063
Not in labour force – Inactifs	0.9779 0.0027	1.0068 0.0027	1.0063 0.0028	1.0073 0.0026	1.0017 0.0026	0.9999 0.0026

Processing error. Processing errors may occur at various stages such as data entry, editing, coding, weighting and tabulation.

During the data capture operation in the regional offices, a sample of LFS forms is verified to monitor and control data entry quality. The method involves the verification (independent re-capture and matching) of a sample of the documents selected from a lot which was processed by a data entry operator. If the number of verified records in the sample that contain one or more data entry errors is not greater than a pre-specified acceptance number, then the lot is accepted. If the number of error records is greater than this number, then the lot is rejected and the remainder of the lot is completely verified. Two measures of the effectiveness of this control process are the **Verification Rate**, which is the percentage of records that were verified for all lots, and the **Average Outgoing Quality**, which is an estimate of the percentage of records which contain data entry errors after the application of the quality control process. Some typical Average Outgoing Quality levels for the eight regional offices and Canada are shown in the table.

A separate, similar sample verification procedure is used to control the quality of the coding operations during the processing stage. Errors in the survey data at the estimation and tabulation stage are controlled through a detailed review of the output, comparison with other data sources and checks for redundancy.

l'enquête pour un certain nombre de fois (premier mois, deuxième, etc.) et l'estimation calculée pour l'échantillon entier. Le tableau qui suit présente la valeur des indices moyens, ainsi que les erreurs types estimées, pour les trois principales caractéristiques de la population active en ce qui concerne la période allant de mars 1985 à février 1988 (Canada).

Le biais de renouvellement de l'EPA a été étudié auparavant par Tessier (1976) et Ghangurde (1982). Ces travaux faisaient suite à l'étude produite par Bailar (1975) portant sur le biais de renouvellement sur l'enquête de la population active des États-Unis (American Current Population Survey).

Indice de renouvellement moyen : mars 1985 – février 1988

Erreur de traitement. Ces erreurs peuvent se produire à diverses étapes du traitement, soit à la saisie, à la vérification, au codage, à la pondération ou à la totalisation des données.

Pendant l'opération de saisie des données aux bureaux régionaux, on prend un échantillon de formules de l'EPA pour surveiller et contrôler la qualité de cette opération. La façon dont on procède consiste à vérifier le travail effectué par un préposé à la saisie des données en prenant dans un lot un échantillon de formules, en introduisant les données à nouveau et en comparant le résultat avec celui obtenu la première fois. Si, dans l'échantillon, le nombre d'enregistrements vérifiés contenant au moins une erreur de saisie ne dépasse pas un seuil établi d'avance, le lot est accepté. Si le nombre d'enregistrements erronés dépasse le seuil, le lot est rejeté et tous les autres enregistrements du lot sont soumis à la vérification. Les deux mesures utilisées pour connaître l'efficacité de ce processus de contrôle sont le **taux de vérification**, qui est le pourcentage d'enregistrements vérifiés sur l'ensemble des lots, et la **qualité moyenne de la sortie (QMS)**, qui est une estimation du pourcentage d'enregistrements contenant encore des erreurs de saisie après le contrôle de la qualité. Le tableau qui suit indique quelques valeurs typiques obtenues pour les huit bureaux régionaux et le Canada.

On procède à une vérification par échantillon distincte mais semblable pour contrôler la qualité des opérations de codage. Pour détecter les erreurs susceptibles de se produire à l'étape de l'estimation et de la totalisation, on procède à un examen soigné du résultat de ces opérations, à une comparaison avec d'autres sources de données et à un contrôle par redondance.

Average Outgoing Quality: July 1988 – December 1988**Qualité moyenne de la sortie : juillet 1988 – décembre 1988**

Regional Office	High	Low	Average
Bureau régional	Maximum	Minimum	Moyenne
St. John's	1.0	0.3	0.6
Halifax	1.2	0.5	0.9
Montreal	0.9	0.4	0.7
Sturgeon Falls	1.3	0.7	1.0
Toronto	1.4	0.6	1.0
Winnipeg	1.0	0.7	0.8
Edmonton	1.3	0.6	1.0
Vancouver	1.3	0.5	0.8
Canada	1.2	0.9	1.0

9.5 Data Collection Quality

The LFS incorporates a number of measures designed for the assessment and control of non-sampling errors that occur in the collection stage of the survey.

Hiring Practices

Prior to being hired as interviewers, applicants are evaluated on such aspects as aptitude and ability to deal with the materials used on the job. Before training starts, new interviewers are sent copies of "Introduction to Interviewing" and "Statistics Canada and You". These booklets introduce the job of LFS interviewer and the organization of Statistics Canada.

Training

Initial training of interviewers is spread over a two-month period. It begins with a three-day classroom training session, during which interviewers receive practical instruction in filling out the different survey and administrative forms. As well, they complete practice exercises, listen to interviews on tape, perform mock interviews and receive guidance in interviewing techniques.

During the interviewers' first survey week they receive two days of on-the-job instruction and one day of this instruction during the second survey. During this time they are accompanied by the senior interviewer, who observes and guides the interviewing as well as demonstrates actual interviews. To keep the interviewers' skills fresh, special training and refresher courses are given at least once a year during a group training session.

The work of the interviewers is assessed as a result of other programs which will be described below. Depending on the interviewer's performance, additional home study and review exercises may be assigned to clarify or reinforce weak points.

9.5 Qualité de la collecte des données

L'EPA prévoit un certain nombre de mesures visant à évaluer les erreurs non dues à l'échantillonnage qui se produisent pendant la collecte des données et à en limiter le nombre.

Recrutement

Avant d'embaucher les candidats à des postes d'intervieweur, on évalue leurs aptitudes et leur capacité à bien remplir les documents de l'enquête. Avant même que la formation ne commence, on leur envoie un exemplaire des documents "Initiation aux interviews" et "Statistique Canada et vous". Ceux-ci présentent le travail des intervieweurs de l'EPA et décrivent l'organisation de Statistique Canada.

Formation

La période initiale de formation des intervieweurs s'étend sur deux mois. Elle commence par un cours de trois jours, en classe, pendant lequel on montre aux intervieweurs qui viennent d'être embauchés comment remplir les formules d'enquête et les documents administratifs. En outre, ces derniers y font des exercices pratiques, écoutent des interviews enregistrées, en simulent eux-mêmes et apprennent les techniques d'interview.

Ensuite, les intervieweurs reçoivent deux jours de formation en milieu de travail pendant la première semaine d'enquête à laquelle ils travaillent et un jour pendant la deuxième. Pour cela, l'intervieweur principal les accompagne : il les observe et leur indique comment mener les interviews et leur donne l'exemple en en menant lui-même. Les intervieweurs participent également à des séances spéciales de formation et de recyclage, en groupe, au moins une fois par an.

Le travail des intervieweurs est évalué dans le cadre d'autres programmes, qui seront décrits plus loin. Selon le rendement de chacun, on détermine s'il y a lieu d'ajouter des cours d'autoformation ou des exercices de révision pour éclaircir certains points ou remédier à des faiblesses.

Observation

The observation program is designed to minimize interviewer error by providing the senior interviewer with the opportunity to observe and assess the performance of the interviewers they supervise and to identify any problems. Each interviewer is observed at least once every 12 months. The scheduling and selection of interviewers to be observed is done by the regional offices, thus ensuring there is no conflict with the re-interview program and no identifiable pattern. The senior interviewer accompanies the interviewer for one day, observing any face-to-face and telephone interviews. A check of cluster listings is performed on a second day. Results of observation are sent to ROs by the senior interviewers and periodic reports are prepared for Head Office. The interviewer receives feedback on his or her performance from the senior interviewer as soon as possible after the observation.

Re-interview

Each month, except for December, a number of dwellings in the LFS sample are chosen for the re-interview program. This sub-sample is selected so that each interviewer will have a part of his or her assignment re-enumerated at least twice each year.

The re-interview is conducted by telephone by RO supervisory staff or senior interviewers in the week following survey week. The re-interviewer conducts a second interview in selected dwellings, making allowance for the fact that the reference week is now two weeks prior to re-interview week.

The re-interview sample is divided into two parts: the reconciled part and the non-reconciled part. In the reconciled part, after the second interview, the senior interviewer compares answers with those that were obtained by the original interviewer, and with the help of the respondent, determines the "correct" answer. In the non-reconciled part, the senior interviewer simply conducts a second independent interview. The reconciled sample data are used to obtain measures of **response bias**, the assumption being that the same error will not be made by both interviewer and re-interviewer and that the reconciliation process will uncover the true response. Data from the non-reconciled part of the re-interview sample are used to estimate **response variance**. The reconciled part also gives the senior interviewer the ability to monitor the interviewer's performance. After re-interview, the senior interviewer discusses discrepancies with the interviewer, offering guidance and identifying the items most susceptible to error.

A measure of bias is the **net difference rate**. For a characteristic such as "employed", it is defined as

$$\frac{[\text{no. of employed in original interview} - \text{no. of employed in re-interview}]/n}{n}$$

Observation

Le programme d'observation vise à réduire au maximum les erreurs que les intervieweurs pourraient commettre en donnant à l'intervieweur principal l'occasion d'observer ceux qui relèvent de lui, d'évaluer leur rendement et de déceler les problèmes. Chaque intervieweur est observé au moins une fois tous les douze mois. C'est au bureau régional qu'on décide qui sera observé et quand, de manière à ce qu'il n'y ait pas de conflit avec le programme de réinterview et qu'on ne puisse pas deviner l'ordre dans lequel le programme d'observation se déroule. L'intervieweur principal accompagne l'intervieweur pendant toute une journée et voit comment se passent les interviews sur place et les interviews téléphoniques. Le deuxième jour, il vérifie les listes de grappe. Il envoie ensuite les résultats de l'observation au BR et rédige des rapports périodiques à l'intention du bureau central. L'intervieweur principal transmet à l'intervieweur le résultat de sa performance aussitôt que possible après l'observation.

Réinterview

Chaque mois, sauf en décembre, on choisit un certain nombre de logements dans l'échantillon de l'EPA pour le programme de réinterview. Ce sous-échantillon est prélevé de manière à ce qu'un partie de la tâche de chaque intervieweur soit réinterviewée au moins deux fois par an.

Les surveillants au BR, ou les intervieweurs principaux, procèdent à la réinterview par téléphone dans le courant de la semaine suivant la semaine d'enquête. Il le font pour les logements choisis en tenant bien compte du fait que la semaine de référence a eu lieu deux semaines auparavant.

L'échantillon constitué pour la réinterview est divisé en deux parties. Pour l'une, l'intervieweur principal compare les réponses obtenues à la seconde interview avec celles fournies la première fois à l'intervieweur et, s'il y a des différences, détermine avec l'aide du répondant quelle est la bonne réponse. Pour l'autre partie de l'échantillon, l'intervieweur principal se contente de mener une interview indépendante. On se sert des données recueillies auprès de la partie de l'échantillon qui a fait l'objet de la comparaison pour obtenir une mesure du **biais de réponse**, l'hypothèse étant que l'intervieweur et l'intervieweur principal ne feraient pas les mêmes erreurs et que la comparaison permet de découvrir les bonnes réponses. On se sert des données recueillies auprès de la partie de l'échantillon qui n'a pas fait l'objet de la comparaison pour estimer la **variance de réponse**. La comparaison des réponses permet aussi à l'intervieweur principal de vérifier le travail des intervieweurs. Après la réinterview, l'intervieweur principal commente, avec les intervieweurs, des divergences obtenues dans les résultats, il leur donne des conseils et leur signale les questions auxquelles on associe le plus grand nombre d'erreurs.

Le **taux d'écart net** mesure le biais. Pour une caractéristique comme "personne occupée", on le définit de la manière suivante :

$$\frac{[\text{nbre de personnes occupées à la première interview} - \text{nbre de personnes occupées à la réinterview}]/n}{n}$$

and is based on the n individuals in the reconciled sample. If the net difference rate is negative, then the LFS estimation procedure is likely to produce underestimates for the characteristic.

The response variance for a characteristic such as "employed" is defined as the number of respondents classified as employed in only one interview, divided by twice the sample size. The **index of inconsistency** is a measure of relative response variance. It is defined as the percentage of total variance accounted for by the response variance, i.e.,

$$\frac{\text{response variance}}{\text{response variance} + \text{sampling variance}} \times 100$$

and is based on the non-reconciled sample.

The following table gives average net difference rates and indices of inconsistency for the period from 1980 to 1986.

% Net Difference Rate and % Index of Inconsistency by Labour Force Status 1980-1986 Averages

	% Net Difference Rate		% Index of Inconsistency	
	Taux d'écart net		Indice de non-cohérence	
Employed – Personnes occupées	-0.28	(-0.5 to 0.1)	5.52	(4 to 7)
Unemployed – Chômeurs	-0.28	(-0.4 to -0.1)	24.07	(20 to 32)
Not in Labour Force – Inactifs	0.56	(0.4 to 0.8)	9.16	(7 to 11)

Note: Numbers in brackets give the range of typical values during the seven-year period.

Nota : Les nombres entre parenthèses indiquent l'intervalle des valeurs les plus courantes pendant la période observée.

An index below 20 percent is considered low, one between 20 percent and 50 percent is moderate, and one greater than 50 percent is high.

Performance Feedback

In addition to the programs described above, monthly reports on performance are provided to the ROs. These deal with costs, edit failure rates and response rates. Senior interviewers maintain regular contact with their interviewers. They bring to the interviewer's attention results of the various performance indicators.

9.6 Operations and Analysis

As indicated in Chapter 7, four committees meet regularly to coordinate the operations of the LFS. The Redesign Committee is active only when the LFS is being redesigned. The other three committees deal with the ongoing operations and evaluation of the survey.

selon les données recueillies auprès de la partie de l'échantillon qui a fait l'objet de la comparaison. Si le taux d'écart net est négatif, il est probable que les estimations produites par la méthode de l'EPA soient inférieures à ce qu'elles devraient être.

La variance de réponse pour une caractéristique comme "personne occupée" est calculée de la manière suivante : nombre de répondants classés dans la catégorie des personnes occupées après une interview seulement divisé par le double de la taille de l'échantillon. L'**indice de non-cohérence** est une mesure de la variance de réponse relative. Elle est définie comme étant le pourcentage de la variance totale représenté par la variance de réponse, c'est-à-dire :

$$\frac{\text{variance de réponse}}{\text{variance de réponse} + \text{variance d'échantillonnage}} \times 100$$

et elle est basée sur la partie de l'échantillon qui n'a pas fait l'objet de la comparaison.

Le tableau suivant donne les taux d'écart net et les indices de non-cohérence pour la période 1980-1986.

Taux d'écart net et indice de non-cohérence selon la situation vis-à-vis de l'activité, moyennes pour 1980-1986

	% Net Difference Rate		% Index of Inconsistency	
	Taux d'écart net		Indice de non-cohérence	
Employed – Personnes occupées	-0.28	(-0.5 to 0.1)	5.52	(4 to 7)
Unemployed – Chômeurs	-0.28	(-0.4 to -0.1)	24.07	(20 to 32)
Not in Labour Force – Inactifs	0.56	(0.4 to 0.8)	9.16	(7 to 11)

On estime qu'un indice inférieur à 20% est faible, qu'un indice entre 20 et 50% est moyen et qu'un indice supérieur à 50% est élevé.

Rétroaction sur le rendement

En plus des programmes décrits ci-dessus, le rendement des intervieweurs fait l'objet de rapports mensuels aux BR. Ces rapports ont trait aux coûts, aux taux de rejet à la vérification et aux taux de réponse. Les intervieweurs principaux sont régulièrement en contact avec leurs intervieweurs, et ils portent à leur attention les résultats des divers indicateurs du rendement.

9.6 Opérations et analyse

Comme nous l'avons vu au chapitre 7, plusieurs comités se réunissent régulièrement pour coordonner les opérations de l'EPA. Le Comité du remaniement est actif uniquement pendant le remaniement de l'EPA. Les trois autres s'occupent des opérations permanentes et de l'évaluation de l'enquête.

Operations Committee

The mandate of the Operations Committee is to review the events of survey months and the circumstances under which the surveys were undertaken, to ensure successful operation and to review and recommend proposals for changes to the operations, which will result in the continued successful conduct of the survey.

Content Consultation Group

The mandate of the Content Consultation Group is to act as the focal point for the resolution of issues concerning the subject matter (labour market concepts) of the questionnaire, interviewer training materials, and reference documents such as manuals and communications on the subject matter of the LFS

Data Quality Committee

The mandate of the Data Quality Committee is to examine and evaluate survey quality on a monthly and longer-term basis, to initiate and review proposals for research and development of methods and procedures affecting data quality and to oversee quality oriented research and development.

To ensure the highest possible quality in the data from the LFS, the various quality indicators described above are monitored regularly by the Data Quality Committee. Each month this committee meets to examine and evaluate monthly survey quality and to advise on any aspect of quality that needs attention. To assist in this evaluation, a monthly operations report is compiled, which includes such quality indicators as non-response, discrepancy rates (errors in data detected during editing), slippage, re-interview non-response, interviewer turnover, and sample verification rates. By monitoring these indicators, immediate feedback to one or more specific LFS operations can be provided to control quality on a survey-to-survey basis. The committee also discusses developments that may affect data quality for the survey just completed or for future surveys. Examples include changes in collection methodology or in the questionnaire, unusual field problems, ongoing programs of testing procedures and methodology, etc.

Another regular and more detailed report, the LFS Quality Report, is also produced. This report is intended to provide an in-depth review of the quality measures associated with the LFS (including measures of sampling error) for a six-month period.

Comité des opérations

Ce comité a pour mandat de passer en revue les événements survenus pendant les mois d'enquête et les circonstances entourant la tenue de l'enquête, de voir à ce que les opérations se déroulent bien, d'examiner les changements proposés et d'en recommander l'adoption, le tout afin que l'enquête continue à atteindre ses objectifs.

Groupe de consultation sur le contenu

Ce comité a pour mandat de coordonner le règlement des problèmes touchant le sujet du questionnaire (les concepts relatifs au marché du travail), les documents pour la formation des intervieweurs ainsi que les documents de référence comme les manuels et enfin les communications sur le sujet dont traite l'EPA.

Comité de la qualité des données

Ce comité a pour mandat d'examiner et d'évaluer la qualité des données d'enquête sur une base mensuelle, de proposer et de réviser des projets de recherche et de développement visant la mise au point de méthodes pouvant influencer sur la qualité des données, et enfin de surveiller la recherche et le développement dans ce domaine.

Pour s'assurer que la qualité des données de l'EPA soit la meilleure possible, le Comité de la qualité des données consulte régulièrement les divers indicateurs de la qualité mentionnés plus haut. Il se réunit tous les mois afin d'examiner et d'évaluer la qualité des données mensuelles et d'émettre son opinion sur tout aspect touchant la qualité des données devant être amélioré. Un rapport mensuel sur les opérations est rédigé pour l'aider dans cette tâche. Ce rapport contient des indicateurs de la qualité comme les taux de non-réponse, de divergence (erreurs dans les données relevées au moment de la vérification), de glissement, de non-réponse à la réinterview, de roulement des intervieweurs et de vérification par échantillon. En suivant de près l'évolution de ces indicateurs, le Comité peut intervenir immédiatement auprès des responsables de la ou des opérations de l'EPA concernées afin que l'on puisse contrôler la qualité des données chaque mois. Le Comité discute également des faits nouveaux susceptibles d'influer la qualité des données qui viennent juste d'être recueillies ou qui le seront à l'avenir, notamment des changements apportés aux méthodes de collecte ou au questionnaire, des problèmes inhabituels sur le terrain, de la mise à l'essai continue de procédures et de méthodes, etc.

Un autre rapport, plus détaillé, est produit régulièrement, le Rapport sur la qualité de l'EPA. Celui-ci contient un examen en profondeur des mesures de la qualité associées à l'EPA (y compris des mesures de l'erreur d'échantillonnage) sur une période de six mois.

Chapter 10

General Use of the Labour Force Survey Vehicle

10.1 Introduction

In addition to supplying the monthly labour market data requirements of Statistics Canada, the LFS is used to collect data for various studies of the general population and for specific subsets of the population. This is achieved through special surveys and supplementary surveys. Special surveys consist of special samples of dwellings chosen from dwellings not sampled for the LFS but selected from areas sampled by the LFS. Supplementary surveys consist of additional questions asked of LFS respondents at the time of the LFS interview, although in some cases the supplementary questionnaires are dropped off at the household and collected later.

Special and supplementary surveys are often sponsored by other government departments. By using the LFS sample, the cost of conducting a large scale survey, whether ad-hoc or periodic, can be greatly reduced. In addition to having the survey vehicle already in place, all the information regularly collected on the Household Record Docket (Form 03) and LFS Questionnaire (Form 05) is automatically available for both screening (i.e., identifying a target population, such as persons in a particular age group) and analysis.

10.2 Special Surveys

Examples of recurring special surveys include the Food Expenditure Survey and the Family Expenditure Survey. The former is a monthly sample survey, usually conducted during even numbered years, in which respondents are required to complete a diary of food expenditures. The latter, occurring in odd numbered years, is conducted two months after the completion of the food survey and consists of a recall of full budget expenditures for the preceding calendar year. Random starts from the same clusters as the LFS are reserved for these surveys. In this way the sample households are different from those used in the LFS although they are selected from already active clusters and PSUs, so that local LFS interviewers can be used. This is the most economical means of selecting a separate sample from the LFS frame.

However, special surveys need not be selected from active LFS clusters and PSUs. Different PSUs can be set aside for the special survey, usually from a part of the frame that is not to be selected for the LFS for

Chapitre 10

Utilisation de l'enquête sur la population active pour la réalisation d'autres enquêtes

10.1 Introduction

Statistique Canada, en plus de se servir de l'EPA pour produire les données mensuelles nécessaires sur le marché du travail, a recours à cette enquête pour recueillir des données sur divers aspects de la population dans son ensemble ainsi que de certains sous-groupes de la population. Les enquêtes spéciales sont menées auprès d'échantillons de logements choisis parmi les logements qui n'ont pas été sélectionnés pour l'EPA mais qui se trouvent dans des secteurs échantillonnés pour l'EPA. Les enquêtes supplémentaires sont menées au moyen de questions rajoutées à celle de l'EPA et posées aux enquêtés au moment où on les interview pour cette enquête, quoiqu'il arrive qu'on laisse au ménage un questionnaire supplémentaire, qu'on revient chercher plus tard.

Les enquêtes spéciales et supplémentaires sont souvent parrainées par d'autres ministères ou organismes publics. Le coût d'une grande enquête, ponctuelle ou périodique, peut être considérablement réduit si l'on se sert de l'échantillon de l'EPA. En plus d'avoir toute la structure en place, on dispose du même coup de tous les renseignements recueillis pour créer le dossier du ménage (formule 03) et remplir le questionnaire de l'EPA (formule 05), lesquels peuvent être utilisés à des fins de présélection (c'est-à-dire pour mettre en évidence la population cible : les personnes appartenant à un groupe en particulier) et d'analyse.

10.2 Enquêtes spéciales

L'enquête sur les dépenses alimentaires et l'enquête sur les dépenses des familles sont deux enquêtes spéciales périodiques. La première est une enquête par sondage mensuelle, qui a généralement lieu les années paires et pour laquelle on demande aux répondants de tenir un journal où ils doivent noter toutes leurs dépenses alimentaires. La deuxième a lieu les années impaires, deux mois après la fin de l'enquête sur les dépenses alimentaires, et elle consiste à demander aux répondants de se rappeler toutes les dépenses budgétaires effectuées l'année précédente. On réserve à ces deux enquêtes des origines choisies au hasard dans les mêmes grappes que celles qui sont échantillonnées pour l'EPA; ainsi, les ménages choisis ne sont pas les mêmes que ceux qui participent à l'EPA, mais ils proviennent de grappes et d'UPE qui sont déjà actives, de sorte qu'on peut prendre des intervieweurs des mêmes régions. Il s'agit là de la façon la plus économique de prélever un échantillon distinct dans la base de l'EPA.

Il n'est toutefois pas nécessaire que l'échantillon des enquêtes spéciales soit prélevé dans les grappes et les UPE actives. Il est possible de réserver à ces enquêtes certaines UPE, qui se trouvent dans une partie de la base de sondage

many years, if ever. An example of such a survey was the sample for the Canada Health Survey. For this survey there was a requirement for clustering of the sample, different from the regular LFS, due to the relatively costly health measures component.

10.3 Supplementary Surveys

Supplementary surveys are directed at more detailed information on some aspects of the labour market that are required only periodically, and a capacity for obtaining data of special interest which are not normally produced as part of the national statistical program of Statistics Canada. As well, the data or the ability to obtain them may not be easily available elsewhere. The latter type of special survey is frequently sponsored by other government departments.

10.3.1 Examples of Supplementary Surveys

Household Surveys Division itself sponsors supplementary surveys dealing with labour market related topics, selected largely on the basis of user interest, such as Annual Work Patterns, Job Opportunities and, during the summer months, the identification of students returning to school in the fall.

Outside sponsorship has included surveys on topics such as Fitness and Physical Recreation, Domestic Travel, Child Health, Work History, Apprenticeship and Smoking Habits. Several design options for supplementary surveys exist, which are outlined below. Examples of of special and supplementary surveys, using the LFS vehicle in 1988, are found in section 10.4. More information on special surveys conducted by Statistics Canada is found in the yearly uncatalogued publication "Overview of Special Surveys".

10.3.2 Sample Selection Options

Each rotation group of the LFS can be used to produce representative statistics at the Canada level and for each of the provinces. Depending on the level of reliability necessary to satisfy statistical requirements, from one to six rotation groups can be used to provide supplementary survey estimates. Normally, however, no more than five rotations are used. The rotation group being introduced into the sample for the first time is avoided because of the time-consuming interview workload already required at the first interview.

If a smaller sample is desired than a whole multiple of a rotation group, the stabilization program is used to reduce the number of selected dwellings, and the dwelling weights are adjusted accordingly.

Within a selected dwelling, the supplementary survey may be directed at all eligible Labour Force Survey respondents or specific individuals. Separate

où l'on ne prélèvera plus d'unités pour l'EPA, ou alors pas avant plusieurs années. L'enquête Santé Canada est une des enquêtes pour lesquelles on a procédé de cette manière. Il fallait constituer les grappes autrement que pour l'EPA en raison du coût élevé de la composante sur les mesures de santé.

10.3 Enquêtes supplémentaires

Les enquêtes supplémentaires sont conçues pour recueillir des données détaillées sur certains aspects du marché du travail dont on n'a besoin que périodiquement et pour produire des données d'intérêt particulier qui ne sont pas normalement produites dans le cadre du programme national de Statistique Canada. En outre, il est possible que personne d'autre n'ait ces données ni la capacité de les produire. Les enquêtes spéciales ayant le second objectif sont souvent parrainées par des ministères ou organismes publics.

10.3.1 Exemples d'enquêtes supplémentaires

La Division des enquêtes-ménages parraine elle-même des enquêtes supplémentaires sur des sujets qui ont trait au marché du travail et qui intéressent les utilisateurs, comme l'enquête sur l'activité annuelle, l'enquête sur les perspectives d'emploi et l'enquête, menée en été, sur les étudiants ayant l'intention de retourner aux études en automne.

Parmi les enquêtes parrainées par des organismes autres que Statistique Canada, mentionnons les enquêtes sur l'exercice physique et les activités physiques et sportives, les voyages des Canadiens, la santé des jeunes, l'activité, les apprentis et les habitudes de fumer. Il existe plusieurs possibilités de plans de sondage pour les enquêtes supplémentaires, et celles-ci sont décrites brièvement plus bas. Des exemples d'enquêtes spéciales et d'enquêtes supplémentaires menées au moyen de l'EPA en 1988 sont donnés à la section 10.4. La publication annuelle "Aperçu des enquêtes spéciales", qui ne figure pas au catalogue, contient plus de détails sur les enquêtes spéciales réalisées par Statistique Canada.

10.3.2 Possibilités de plans de sondage

Chaque groupe de renouvellement de l'EPA est un échantillon indépendant à partir duquel il est possible de produire des statistiques représentatives à l'échelle du Canada et de chaque province. Selon le degré de fiabilité qu'il faut atteindre pour satisfaire aux exigences statistiques, on peut prendre d'un à six groupes de renouvellement pour produire les estimations des enquêtes spéciales. Cependant, en général, on n'en prend pas plus de cinq. On évite le groupe de renouvellement qui vient d'être introduit dans l'échantillon parce que la première interview de l'EPA prend déjà beaucoup de temps.

Si on veut un échantillon plus petit que celui qu'on obtiendrait en prenant un multiple entier du nombre de groupes de renouvellement, on a recours au programme de stabilisation pour réduire le nombre de logements choisis et on modifie le poids des logements en conséquence.

Au niveau des logements échantillonnés, l'enquête supplémentaire peut s'adresser à tous les membres admissibles selon les critères de l'enquête sur la population active

individual respondents may be selected from within selected dwellings through random selection or by screening for respondents with specific demographic or labour force characteristics from the LFS documents or through special questions.

10.3.3 Data Collection Methodologies

Depending on the length and complexity of the survey, a number of different data collection methodologies can be employed. The traditional method is a personal interview completed at the time of the regular LFS enumeration. Generally, one page of additional questions can be accommodated using this approach. These single page supplements (Form 06) can accommodate from 15 to 25 questions which the interviewer will ask after completing the LFS Questionnaires for all household members.

When supplementary surveys are long and complex or require self-enumeration (i.e., when proxy reporting is not appropriate), a separate questionnaire may be distributed by mail or left with respondents at the time of the regular LFS interview. For telephone respondent households, arrangements are made for the enumerator to deliver or mail out the documents. Documents are either picked up by enumerators, at which time they may or may not offer assistance in completing forms, or they are returned by mail.

The Household Record Docket may be used for recording the answers to about five very brief and straightforward questions. This area may be used for supplementary data or for recording results of screening questions.

10.3.4 Acceptability Criteria for Supplementary Surveys

The effect of a supplementary survey on respondents is a primary concern. Topics or questions that are likely to be unacceptable to respondents, or to influence in some way responses obtained for the LFS in the following month, are avoided. Supplements must comply with Statistics Canada criteria for reliability of data and confidentiality of responses. In general, supplements appear to be well-received; they increase interviewing time, but on the other hand, also add variety to the experience of being included in the LFS sample for six months.

10.4 Examples

The following table lists the special and supplementary surveys which were conducted in the 1988 calendar year. Indicated are the name of the survey, the month it was conducted, the method of collection (personal interview, telephone interview, self enumeration by mail) and the number of rotation groups that were used.

ou à certaines personnes en particulier. Il s'agit alors soit de prélever un échantillon aléatoire d'individus à l'intérieur des ménages choisis soit de faire une présélection en consultant les documents de l'EPA ou en posant certaines questions afin de retenir les répondants ayant certaines caractéristiques du point de vue démographique ou du point de vue de l'activité.

10.3.3 Méthodes de collecte des données

Il est possible d'utiliser différentes méthodes de collecte des données, dont le choix dépend de la durée et de la complexité de l'enquête. La méthode habituelle consiste à donner des interviews sur place en même temps que celles de l'EPA. En général, on peut ajouter une page de questions si l'on utilise cette méthode. Cette page supplémentaire (formule 06) peut contenir de 15 à 25 questions, que l'intervieweur peut poser à tous les membres du ménage après celles de l'EPA.

Lorsqu'une enquête supplémentaire est longue et complexe ou qu'elle exige que les enquêtés répondent eux-mêmes aux questions (c'est-à-dire qu'on n'accepte pas de réponses par personne interposée), on peut envoyer un questionnaire distinct par la poste ou le remettre aux répondants au moment de l'interview de l'EPA. Dans le cas des ménages qui font l'objet d'une interview téléphonique, on prend les dispositions nécessaires pour que ce soit l'intervieweur qui poste ou dépose le questionnaire. Celui-ci est soit repris par l'intervieweur, qui peut alors offrir de l'aide aux répondants s'ils ont du mal à le remplir tout seuls, soit renvoyé par la poste.

On peut noter dans le dossier du ménage les réponses à environ cinq questions très brèves et directes. Il peut s'agir de questions par lesquelles on cherche à obtenir des données supplémentaires ou de questions de sélection.

10.3.4 Critères d'admissibilité relatifs aux enquêtes supplémentaires

L'effet qu'une enquête supplémentaire est susceptible d'avoir sur les répondants est une préoccupation primordiale. On évite les sujets ou les questions que ces derniers pourraient trouver inacceptables ou qui pourraient influencer le déroulement des interviews de l'EPA le mois suivant. Par ailleurs, les enquêtes supplémentaires doivent se conformer aux critères de Statistique Canada en matière de fiabilité et de confidentialité des données. Dans l'ensemble, elles semblent être bien accueillies; il est vrai qu'elles prolongent l'interview, mais en revanche elles introduisent un élément de variété apprécié par les répondants qui participent à l'EPA pendant six mois.

10.4 Exemples

Le tableau qui suit contient la liste des enquêtes supplémentaires et spéciales menées pendant l'année civile 1988. Il indique le nom des enquêtes, le mois où elles ont eu lieu, la méthode de collecte de données utilisée (interview sur place, interview téléphonique ou questionnaire envoyé par la poste) et le nombre de groupes de renouvellement utilisés.

1988 Special and Supplementary Surveys

Enquêtes spéciales et supplémentaires – 1988

Survey	Month	Method	Size
Enquête	Mois	Méthode	Taille
Labour Market Activity – Activité	Jan. – Janv	Personal & Telephone – Sur place/téléphone	5 Rot Gps – 5 gr. de renouv
Importance of Wildlife to Canadians – Importance de la faune	Feb. – Fév	Mailout/Mailback – Quest. par la poste	5 Rot Gps – 5 gr. de renouv.
Absence from Work – Absence du travail	Feb. – Fév	Personal & Telephone – Sur place/téléphone	5 Rot Gps – 5 gr. de renouv
Shelter Cost – Frais de logement	Mar. – Mars	Personal & Telephone – Sur place/téléphone	5 Rot Gps – 5 gr. de renouv
Job Opportunities – Perspectives d'emploi	Mar. – Mars	Personal & Telephone – Sur place/téléphone	6 Rot Gps – 6 gr. de renouv
Consumer Finances – Finances des consommateurs	Apr. – Avr.	Mail & Telephone – Poste/téléphone	5 Rot Gps – 5 gr. de renouv.
Travel – Voyages	Apr. – Avr	Personal & Telephone – Sur place/téléphone	1 Rot Gp – 1 gr. de renouv
Veterans – Anciens combattants	May – Mai	Personal & Telephone – Sur place/téléphone	4 Rot Gps – 4 gr. de renouv.
Household Facilities and Equipment – Equipement ménager	May – Mai	Personal & Telephone – Sur place/téléphone	4 Rot Gps – 4 gr. de renouv.
Travel – Voyages	Jul. – Juill.	Personal & Telephone – Sur place/téléphone	3 Rot Gps in 5 Provs 1 Rot Gp in 4 Provs – 3 gr. de renouv. dans 5 prov 1 gr. de renouv. dans 5 prov
National Child Care – Garde des enfants	Sept.	Personal & Telephone – Sur place/téléphone	13 Rot Gps – 13 gr. de renouv
Travel – Voyages	Oct.	Personal & Telephone – Sur place/téléphone	3 Rot Gps in 6 Provs 1 Rot Gp in 4 Provs – 3 gr. de renouv. dans 6 prov 1 gr. de renouv. dans 4 prov
Apprentices and Journeymen – Apprentis et manoeuvres	Oct	Personal & Telephone – Sur place/téléphone	5 Rot Gps in 1 Prov – 5 gr. de renouv
Health and Employment – Santé et état d'emploi	Nov	Personal & Telephone – Sur place/téléphone	5 Rot Gps – 5 gr. de renouv.
Telephone Services – Service téléphonique	Nov.	Personal & Telephone – Sur place/téléphone	6 Rot Gps – 6 gr. de renouv.

Chapter 11

Developmental Projects

11.1 Introduction

While decennial redesigns offer the principal opportunity to implement the major changes in the survey instrument, some improvements can be implemented in between redesigns. Traditionally, therefore, developmental work has been carried out to identify improvements in the survey methodology as a continuous feature of the LFS.

Following the sample redesign during the early 1960s, several developmental studies were initiated that had a large influence on the survey methodology introduced during the revision and redesign of the 1970s. Studies included detailed cost and variance component analyses, the effects of proxy-vs. self-response, and data collection research that led to introduction of telephone interviewing for non-first month households in Self Representing areas.

Studies conducted from 1976 to 1981 included the development of sample updating methodology for SR areas, sample size stabilization methodology, and methods for increasing or decreasing the sample size. All these developments were implemented during this period. During this period, as well, research and evaluation of alternative small area estimation techniques were continued. The cost and variance models were reformulated and used to study alternative sample designs, with a view to optimizing the number of stages of sampling, and the allocation of sample to stages. These studies were influential in identifying design changes introduced during the 1981 post-censal redesign.

Following the introduction of the 1981 post-censal redesign, new developmental studies began. As was the case in earlier decades, some of these have already led to the implementation of improved survey methods. Examples of this are the regression method of estimation introduced in 1988, and the introduction of some new telephone collection procedures.

In this chapter, the major developmental projects leading up to the 1991 post-censal redesign are described. They include small area estimation, composite estimation, gross flows, longitudinal data, variance studies, telephone and CATI development, and Address Register development.

Chapitre 11

Projets de développement

11.1 Introduction

Bien que les remaniements décennaux de l'EPA soient l'occasion privilégiée pour la mise en oeuvre des changements importants apportés à l'enquête, il est possible d'introduire des améliorations entre ces remaniements. Il y a donc toujours des travaux de développement en cours, dont le but est de découvrir des façons de perfectionner la méthodologie de l'EPA.

A la suite du remaniement du plan de sondage qui a eu lieu au début des années 60, on a entrepris un certain nombre d'études de développement qui ont eu une grande influence sur la méthodologie introduite lors de la révision et du remaniement des années 70. Parmi celles-ci, on comptait des analyses détaillées des coûts et des composantes de la variance, une étude de l'effet des réponses par procuration par opposition aux réponses fournies en personne, et des travaux de recherche sur la collecte des données qui ont abouti à l'adoption de l'interview téléphonique auprès des ménages dans les secteurs autoreprésentatifs qui ne participent pas à l'enquête pour la première fois.

Les études réalisées entre 1976 et 1981 ont conduit à l'élaboration d'une méthode de mise à jour de l'échantillon pour les secteurs AR, d'une méthode de stabilisation de la taille de l'échantillon et de méthodes d'élargissement et de rétrécissement de la taille de l'échantillon. Toutes ces méthodes ont été mises en application aussitôt élaborées. On a continué pendant la même période à chercher et à évaluer d'autres techniques d'estimation pour les petites régions et on a reformulé les modèles de coût et de variance, qu'on a utilisés pour étudier d'autres plans de sondage éventuels qui permettraient d'optimiser le nombre de degrés d'échantillonnage et la répartition de l'échantillon entre les degrés. On s'est basé sur le résultat de ces études pour décider des changements à apporter au plan de sondage après le recensement de 1981.

De nouvelles études de développement ont été entreprises après le remaniement du plan de sondage qui a suivi le recensement de 1981. Comme par le passé, certaines de ces études ont déjà mené à l'adoption de meilleures techniques d'enquête. On peut citer par exemple la méthode d'estimation par régression introduite en 1988 et les nouvelles méthodes de collecte de données par téléphone.

Nous allons décrire dans ce chapitre les principaux projets de développement en cours en prévision du remaniement du plan de sondage qui aura lieu après le recensement de 1991. Ils portent notamment sur l'estimation pour les petites régions, l'estimation composite, les flux bruts, les données longitudinales, les études de la variance, les nouvelles façons d'utiliser l'interview téléphonique et de l'interview téléphonique assistée par ordinateur et enfin le registre d'adresses.

11.2 Small Area Estimation

A major focus of the post-1981 censal redesign of the LFS was on improved subprovincial data. Two principal measures were taken in this respect. First, as described in Chapter 5, major changes were made to the sample allocation scheme. Whereas in previous designs, the Non Self-Representing and Self-Representing areas were sampled at uniform rates within each province, in the redesign, variable sampling rates were adopted so as to achieve pre-specified levels of reliability for monthly estimates for Census Metropolitan Areas and Economic Regions and quarterly estimates for Census Agglomeration areas.

Second, a flexible small area estimation capacity was introduced to meet the demand for survey estimates for other subprovincial areas, which could not be explicitly taken into account when designing the survey. We refer to such areas as unplanned domains.

Boundaries of the unplanned domains will typically cross the survey design strata. Given the clustered nature of the LFS sample, particularly in NSR areas, the unplanned domains can range from being severely under-sampled to severely over-sampled; this is relative to rates at which they would have been sampled had they been design strata.

Following extensive research and evaluation of alternative estimation methodologies for these unplanned domains, the sample size dependent estimator proposed by Drew, Singh and Choudhry (1982) was implemented. As the name suggests, this particular estimator depends on the outcome of the given sample. If the sample size within the domain is reasonably large, then the sample size dependent estimator is based entirely on the sample falling in the domain. If the sample size is not sufficiently large, the estimate for the small area is constructed as a linear combination of an estimate based on sample falling in the domain, and an estimate based on sample from those strata which are partly in the domain. In such cases, the reliance on sample falling outside the domain increases as the amount of the sample falling inside the domain decreases. The sample size dependent estimator is obtained separately for each of the area types identified in the design of the survey such as SR, NSR- urban, NSR-rural, and so on. The overall estimate from the sample falling outside the domain will be less than a certain prespecified maximum acceptable value – otherwise the domains are combined.

If independent population estimates for the small areas are also available, the LFS estimates are benchmarked to these external population estimates. There is significant improvement in the reliability of the LFS estimates due to population adjustment at the small area level.

11.2 Estimation pour les petites régions

Un des objectifs principaux du remaniement de l'EPA qui a eu lieu après le recensement de 1981 était l'amélioration des données infraprovinciales. À cet égard, on a pris principalement deux mesures. Premièrement, comme nous l'avons vu au chapitre 5, le mode de répartition de l'échantillon a été grandement modifié. Alors qu'auparavant on avait utilisé des fractions de sondage uniformes pour les secteurs non autoreprésentatifs et les secteurs autoreprésentatifs de chaque province, à l'occasion du remaniement, on a adopté des fractions de sondage variables afin que les estimations mensuelles relatives aux régions métropolitaines de recensement et aux régions économiques et les estimations trimestrielles relatives aux agglomérations de recensement atteignent certains niveaux de fiabilité pré-établis.

Deuxièmement, on a introduit une méthode d'estimation souple pour les petites régions afin de pouvoir répondre à la demande d'estimations concernant les autres régions infraprovinciales dont il n'avait pas été tenu compte de façon explicite au moment de l'élaboration du plan de sondage. Nous appelons ces régions des domaines non prévus.

Les domaines non prévus ont tendance à chevaucher les strates définies par le plan de sondage. Étant donné que l'échantillon de l'EPA est réparti en grappes, particulièrement dans les secteurs NAR, les domaines non prévus peuvent être sérieusement sous-échantillonnés comme ils peuvent être sérieusement suréchantillonnés par rapport au taux de sondage qui aurait été le leur s'il s'agissait de strates définies par le plan.

À la suite de recherches poussées et de l'évaluation des autres méthodes d'estimation qu'on aurait pu appliquer à ces domaines non prévus, on a adopté l'estimateur qui dépend de la taille de l'échantillon proposé par Drew, Singh et Choudhry (1982). Comme son nom l'indique, cet estimateur dépend de la taille effective de l'échantillon. Si celle-ci est raisonnablement grande pour le domaine, l'estimateur est fondé entièrement sur l'échantillon qui se trouve à l'intérieur du domaine. Si elle n'est pas assez grande, l'estimateur pour la petite région est construit de manière à être une combinaison linéaire d'un estimateur fondé sur l'échantillon qui tombe dans le domaine et d'un estimateur fondé sur un échantillon tiré dans les strates qui sont en partie dans le domaine. Dans ces cas-là, le recours à l'échantillon à l'extérieur du domaine augmente au fur et à mesure que diminue la taille de l'échantillon à l'intérieur du domaine. L'estimateur lié à la taille de l'échantillon est calculé séparément pour chacun des genres de secteur définis par le plan de sondage de l'enquête, soit les SA, les parties urbaines des SNAR, les parties rurales des SNAR, etc. L'estimation globale pour l'échantillon à l'extérieur du domaine doit être inférieure à une valeur maximum acceptable établie d'avance, sans quoi les domaines sont combinés.

Si on a, en plus, accès à des chiffres de population indépendants relativement aux petites régions, on s'en sert comme données de référence pour les estimations de l'EPA. La fiabilité des données de l'EPA est augmentée de façon significative par les corrections apportées aux chiffres de population au niveau des petites régions.

Finally, in the case of an exhaustive set of domains covering an entire province the estimates are also raked to achieve consistency by labour force status (employed, unemployed, and not in labour force) at the province level at the same time as assuring consistency for the total population at the small area level. If the external population estimates for the small area were not available, then the LFS estimates are simply ratio adjusted to achieve consistency by labour force status at the province level.

Unplanned domains include Census Divisions, for which three-year moving average estimates are produced each year, and Canada Employment Centres and Federal Electoral Districts, for which annual estimates have been produced.

In the post-redesign period, research has continued into improved small area estimation. Choudhry and Hidirolou (1988), and Choudhry and Rao (1989) have investigated time series regression approaches which in addition to using the survey data, make use of external data such as counts of Unemployment Insurance beneficiaries. The more sophisticated methods currently under investigation have potential for improving estimation for planned as well as unplanned domains.

11.3 Composite Estimation

For continuous surveys, generally both estimates of level and estimates of change are important. The requirement for these two types of estimates can be taken into account both in the sample design and in the estimation methods used for the survey.

In the Labour Force Survey, the sample design reflects the need for both types of estimate. For example, the sample was allocated to provide reliable monthly estimates of level for the provinces and for subprovincial regions. Similarly, the rotating panel design of the LFS, in which five-sixths of the sample is common from one month to the next, results in efficient estimates of change. This is because most characteristics of individuals and households are positively correlated over time. Current LFS estimation procedures, on the other hand, do not explicitly take into account the requirement for the two types of estimates. Estimates of level are constructed independently for each survey occasion, following the procedures outlined in Chapter 8, and estimates of change are derived simply as the differences between respective estimates of level. Specifically, there is no exploitation of the overlapping nature of the sample design in order to improve estimates of change, or, for that matter, estimates of level.

Enfin, dans le cas d'un ensemble exhaustif de domaines recouvrant une province entière, les estimations sont également sournies à une méthode de balayage qui vise à les rendre conformes aux estimations selon la situation vis-à-vis de l'activité (personnes occupées, en chômage et inactives) au niveau provincial et à faire en sorte que les données relatives à la population totale concordent avec celles sur les petites régions. Si on n'a pas de chiffres de population pour les petites régions en provenance d'une autre source, on redresse simplement les estimations de l'EPA par la méthode du quotient de manière qu'elles soient conformes aux estimations selon la situation vis-à-vis de l'activité au niveau provincial.

Les domaines non prévus comprennent les divisions de recensement, pour lesquelles on produit chaque année des estimations de moyennes mobiles de trois ans, ainsi que les Centres d'emploi du Canada et les circonscriptions électorales fédérales, pour lesquelles on produit des estimations annuelles.

Les travaux de recherche se sont poursuivis après le remaniement du plan de sondage en vue de l'amélioration de l'estimation pour les petites régions. Choudhry et Hidirolou (1988) et Choudhry et Rao (1989) se sont penchés sur les méthodes de régression appliquées à des séries chronologiques, qui utilisent non seulement les données d'enquête mais aussi d'autres données comme le nombre de bénéficiaires de l'assurance-chômage. Les méthodes sophistiquées qui sont actuellement à l'étude pourraient éventuellement améliorer l'estimation tant pour les domaines prévus que pour les domaines non prévus.

11.3 Estimation composite

Avec les enquêtes permanentes, il est généralement important de produire à la fois des estimations de niveau et des estimations de la variation. Il est possible de tenir compte des exigences relatives à ces deux sortes d'estimations soit dans le plan de sondage soit dans les méthodes d'estimation choisies pour l'enquête.

Pour ce qui est de l'enquête sur la population active, c'est dans le plan de sondage que les deux sortes d'estimation sont prévues. Ainsi, l'échantillon est réparti de manière à ce qu'on puisse produire des estimations mensuelles des niveaux qui soient fiables pour les provinces et pour les régions infraprovinciales. Le mode de renouvellement de l'échantillon, selon lequel les cinq sixièmes de l'échantillon ne changent pas d'un mois à l'autre, permet aussi d'obtenir des estimations efficaces de la variation, puisque les caractéristiques des particuliers et des ménages sont en corrélation directe dans le temps. En revanche, les méthodes d'estimation de l'EPA qui sont utilisées en ce moment ne tiennent pas compte explicitement des exigences relatives aux deux sortes d'estimations. Les estimations de niveau sont calculées indépendamment à chaque passage de l'enquête, de la manière décrite au chapitre 8, et les estimations de la variation sont simplement obtenues en faisant la différence entre les estimations de niveau correspondantes. Autrement dit, on n'exploite pas le fait qu'il y a un chevauchement de l'échantillon en vertu du plan de sondage pour améliorer les estimations de la variation ni, de fait, les estimations des niveaux.

During the research phase of the post-1981 censal redesign, composite estimation methods were investigated as a means of taking the rotating panel nature of the design into account in the estimation procedure. For month m , the composite estimate y'_m is defined by:

$$y'_m = (1 - K)y_m + K(y'_{m-1} + d_{m-1,m}), 0 < k < 1, (1)$$

where y_m is the current estimate, and $d_{m-1,m}$ is the estimate of change based on the common part of the sample in months m and $m-1$, and K is an arbitrary constant whose value is typically chosen based on efficiency considerations.

Hence the composite estimate of level for month m is a weighted average of (i) the usual cross-sectional estimate for month m , and (ii) the composite estimate for month $m-1$ plus the estimate of change from month $m-1$ to m based on the common sample.

The study of composite estimation was undertaken by Kumar and Lee (1983). Their key findings, and conclusions on areas requiring further investigation, are described below.

Examination of the composite estimate defined in (1) indicated that the optimal value of K varies depending on the characteristic and on the type of estimate (change or level). The use of the optimal values of K for the different estimates would result therefore in consistency problems. For example, the optimal estimate of persons in the labour force would not equal the sum of the optimal estimates of persons employed and of persons unemployed. A common value of $K = 0.4$ was found, though, to be a reasonable compromise.

The study of composite estimation was extended to consider a broader class of estimators, referred to in the literature as 'AK' composite estimators. These were investigated primarily to handle complications arising from the fact that estimates based on the panel or rotation group in the sample for the first time differ in expected value from estimates based on the remaining panels. This phenomenon, known as rotation group bias, was reported by Bailer (1975). The AK-composite estimate is defined as follows:

$$y''_m = (1 - K + A) \frac{y_{m,1}}{6} + (1 - K - A/5) \sum_{j=1}^6 \frac{y_{m,j}}{6} + K(Y''_{m-1} + d_{m,m-1}), (2)$$

where $y_{m,j}$ is the estimate for the j -th panel for month m , and where A is another constant.

Pendant la période de recherche qui a précédé le remaniement du plan de sondage du début des années 80, on a étudié des méthodes d'estimation composite pour voir s'il serait possible de tenir compte, dans la méthode d'estimation, du fait que le plan de sondage est un plan avec renouvellement de l'échantillon. Pour le mois m , l'estimation composite y'_m est défini par l'équation :

$$y'_m = (1 - K)y_m + K(y'_{m-1} + d_{m-1,m}), 0 < k < 1, (1)$$

où y_m est l'estimation courante, d_{m-1} est l'estimation de la variante fondée sur la partie de l'échantillon commune aux mois m et $m-1$ et K est une constante arbitraire dont la valeur est généralement choisie selon des critères d'efficacité.

L'estimation composite du niveau pour le mois m est donc une moyenne pondérée (i) de l'estimation transversale normale pour le mois m et (ii) de l'estimation composite pour le mois $m-1$ plus l'estimation de la variation du mois m par rapport au mois $m-1$ selon la partie de l'échantillon commune au deux.

L'étude de l'estimation composite a été entreprise par Kumar et Lee (1983). Les principaux résultats et les conclusions concernant les domaines nécessitant des recherches plus poussées sont décrits ci-après.

Il est ressorti de l'examen de l'estimation composite définie par l'équation (1) que la valeur optimale de K varie selon la caractéristique et la sorte d'estimation (variation ou niveau). Si on utilisait les valeurs optimales pour les différentes estimations, on aurait des problèmes de cohérence. L'estimation optimale du nombre de personnes dans la population active, par exemple, ne serait pas égale à la somme de l'estimation optimale du nombre de personnes occupées et du nombre de personnes en chômage. On a toutefois trouvé qu'une valeur de $K = 0.4$ pour toutes les estimations était un compromis acceptable.

On a élargi le champ de l'étude pour examiner une catégorie plus grande d'estimateurs, connus dans les ouvrages spécialisés sous le nom d'estimateurs composites 'AK'. Le but principal était de résoudre les complications qui surviennent du fait que la valeur prévue des estimations fondées sur le groupe de renouvellement inclus dans l'échantillon pour la première fois est différente de celle des estimations fondées sur les autres groupes. Ce phénomène, qu'on appelle le biais de renouvellement, a été observé par Bailer (1975). L'estimation composite AK est définie par l'équation :

$$y''_m = (1 - K + A) \frac{y_{m,1}}{6} + (1 - K - A/5) \sum_{j=1}^6 \frac{y_{m,j}}{6} + K(Y''_{m-1} + d_{m,m-1}), (2)$$

où $y_{m,j}$ est l'estimation pour le j -ième groupe pour le mois m et où A est une autre constante.

This composite estimate was initially proposed by Gurney and Daly (1955). Note that y''_m reduces to y'_m when $A=0$. As it defines a bigger class, the AK -composite estimate can improve the variance efficiency over the K -composite estimate defined by (1).

If it is assumed that the normal cross-sectional estimate y_m is unbiased, and if mean square error is the criterion for efficiency, then the K -composite estimate performs poorly, since in effect it results in unequal weighting of the rotation groups. However, by choosing appropriate values for A and K , the bias with respect to y_m can be controlled. Study findings were that the AK -composite estimate is superior to the ratio and K -composite estimates with respect to a mean square error criterion.

The initial studies on composite estimation were positive, but they pointed to areas requiring further research and evaluation. Hence the decision was taken not to implement composite estimation coincident with the 1981 post-censal redesign of the sample, and to defer its further study until the post-redesign period. Some aspects where further investigation is planned preparatory to the next redesign include the following: the impact of changing the estimation method for the cross-sectional estimate from the raking ratio to the regression method on the choice of the parameters A and K ; the stability of efficiency gains for choices of A and K over longer periods of time; the effect of composite estimates on seasonally adjusted series; and implementation options, including, for operational convenience, the actual derivation of composite estimates only for provincial estimates of unemployed, employed and not in the labour force, and using these as control totals in the regression estimate.

11.4 Gross Flows

For panel survey data, it is of common interest to estimate gross flows (i.e., frequencies of individuals making transitions between categories or continuing in the same category from one time point to another) for a characteristic that is categorical in nature. For example, in the LFS, the characteristic of interest might be the labour force status for each individual, which is reported as either E (Employed), U (Unemployed), or N (Not in the labour force). The (E, U, N) table of estimated gross flows for time points $t-1$ and t is important for better understanding labour market dynamics and their impact on individual members of the population, than is possible from the analysis of the net flows from the cross-sectional data. Of special interest are the off-diagonal cells indicating monthly flows from one labour force status to another.

A major problem in estimating gross flows other than the usual problem of non-response arises due to the presence of response or classification errors. It is commonly believed that biases in the stock levels (that is, the marginal frequencies in the gross flow table) are negligible, but the gross flows or interior frequencies of the table may be considerably biased. In particular,

Gurney et Daly (1955) ont été les premiers à proposer cette estimation composite. Remarquons que y''_m devient y'_m lorsque $A=0$. Comme l'estimation composite AK représente une plus grande catégorie, elle constitue une amélioration par rapport à l'estimation composite K définie par l'équation (1) du point de vue de l'efficacité de la variance.

Si l'on suppose que l'estimation transversale normale y_m n'est pas biaisée et si l'erreur quadratique moyenne est le critère d'efficacité, l'estimateur composite K est peu efficace, car il en découle une pondération inégale pour les groupes de renouvellement. Toutefois, on peut réduire le biais de y_m en attribuant à A et à K des valeurs appropriées. Selon les résultats de l'étude, si l'on prend pour critère l'erreur quadratique moyenne, l'estimation composite AK est supérieure à l'estimation par quotient et à l'estimation composite K .

Les premières études réalisées à propos de l'estimation composite étaient positives, mais elles ont mis en évidence des domaines sur lesquels il fallait pousser les recherches et les évaluations. Il a donc été décidé de ne pas adopter l'estimation composite au moment du remaniement du plan de sondage du début des années 80 et de remettre à après la période de remaniement la poursuite des travaux d'étude d'y rapportant. Parmi les questions qui vont faire l'objet d'autres recherches en prévision du prochain remaniement du plan de sondage, mentionnons les conséquences qu'aura sur le choix des paramètres A et K le fait de passer de la méthode du quotient à la méthode de régression pour les estimations transversales; la stabilité à plus long terme des gains d'efficacité pour des valeurs choisies de A et de K ; l'effet des estimations composites sur les séries désaisonnalisées; la mise en œuvre d'options dont, pour des raisons d'ordre pratique, le calcul d'estimations composites seulement pour les estimations provinciales du nombre de personnes en chômage, occupées et inactives et l'utilisation de ces estimations comme totaux de contrôle dans l'estimation par régression.

11.4 Flux bruts

Lorsqu'on dispose de données tirées d'enquêtes par panel, il est toujours intéressant d'estimer les flux bruts (c'est-à-dire la fréquence à laquelle les personnes passent d'une catégorie à l'autre ou restent dans la même catégorie entre deux moments donnés) pour une caractéristique qualitative. Dans l'EPA, par exemple, la caractéristique à l'étude pourrait être la situation de chaque personne vis-à-vis de l'activité, déclarée comme étant O (occupé), C (en chômage) ou I (inactif). Le tableau (O, C, I) des flux bruts estimés pour les moments t et $t-1$ est important si l'on veut mieux comprendre la dynamique du marché du travail et ses effets sur les membres de la population qu'on ne peut le faire en analysant les flux nets dérivés des données transversales. Les cases qui ne sont pas sur la diagonales et qui indiquent les flux mensuels d'une situation vis-à-vis de l'activité à une autre sont particulièrement intéressantes.

Un gros problème, autre que celui de la non-réponse, auquel on se heurte lorsqu'on estime les flux bruts est dû à la présence d'erreurs de réponse ou de classement. Beaucoup croient que les biais dont sont entachées les estimations de stocks (c'est-à-dire la fréquence marginale dans le tableau de flux bruts) sont négligeables, mais il se peut que les flux bruts ou fréquences intérieures soient considérablement biaisés. Il

there may be serious upward biases in the off-diagonal cell flows. This follows since even small classification errors for the large diagonal cells (which correspond to no change in status) translate into large relative errors for the much smaller off-diagonal cells (which correspond to changes in status). The factors that could contribute to classification errors include proxy response, a misunderstanding of questions asked in the interview, coding errors, and interviewer errors.

A popular method for correcting classification errors is based on the availability of reinterview data for estimating error rates under the assumption of independent classification errors; see, for example, papers by Abowd and Zellner (1985), and Chua and Fuller (1987). By independent classification errors we mean that the probability of the observed classification being in error at time t depends only on the true classification at time t , and not on the true or observed classification at time $t-1$. On the other hand, if the response errors were serially dependent, say, positively correlated, this would tend to decrease the number of reported changes and increase the number of reported continuations of the previous category. Hence the adjustment required to correct for classification errors would be smaller under positively correlated classification errors than under independent classification errors.

Some important efforts were made at Statistics Canada on addressing the problem of assessing classification bias in gross flows. An empirical evaluation of the methods based on the independent classification errors assumption was made by Lemaître (1988) which showed that these methods might be overcorrecting for response bias. Gentleman (1988) discussed implications of some commonly made assumptions such as unbiased stock estimates and independent classification errors, and concluded that these assumptions led to results that were not supported by the empirical data used in her study. Singh and Lemaître (1989) attempted to avoid the assumption of independent classification errors by using an ε -contamination model for response error. An alternative source, namely administrative data on unemployment insurance (UI) beneficiaries, was used instead of reinterview data for estimating the model parameter ε analogous to response error probabilities. It should be noted that the independent errors assumption would not have been needed if it were possible to carry out double re-interviews, (that is, re-interviewing the same respondents in two consecutive months), but this approach was not pursued due to operational and response burden considerations.

In the ε -contamination model, the response errors are specified by means of a contamination distribution assumed for individuals prone to response error. This distribution is estimated under the assumptions that stock estimates are unbiased and that responses for the two time points are independent only for those individuals who are prone to error. The mixing parameter ε (roughly the fraction of individuals in the population who are prone to response error) provides a

se peut en particulier qu'il y ait d'importants biais par excès dans les flux des cases qui ne sont pas sur la diagonale, car même de petites erreurs de classement pour les cases diagonales dont l'effectif est élevé (qui correspondent à aucun changement de situation) se traduisent par de grosses erreurs relatives pour les cases qui ne sont pas sur la diagonale et dont l'effectif est beaucoup moins élevé (qui correspondent à un changement de situation). Les facteurs susceptibles de causer des erreurs de classement sont les réponses par procuration, le fait de ne pas comprendre ou de mal comprendre les questions posées pendant l'interview, les erreurs de codage et les erreurs d'interview.

Il existe une méthode très utilisée pour la correction des erreurs de classement qui consiste à estimer les taux d'erreur à l'aide des données provenant des réinterviews et repose sur l'hypothèse selon laquelle les erreurs de classement sont indépendantes; voir par exemple l'article de Abowd et Zellner (1985) et celui de Chua et Fuller (1987). Par erreurs de classement indépendantes, on veut dire que la probabilité que le classement observé soit erroné au moment t dépend seulement du classement réel au moment t et non du classement réel ou observé au moment $t-1$. Si, en revanche, les erreurs de réponse étaient liées (si elles étaient en corrélation directe par exemple), le nombre de changements déclarés serait inférieur et le nombre de réponses identiques aux réponses données antérieurement serait supérieur. La modification nécessaire pour corriger l'erreur de classement serait moins grande dans le cas des erreurs de classement en corrélation directe que dans le cas des erreurs de classement indépendantes.

De gros efforts ont été faits à Statistique Canada pour régler le problème que pose l'évaluation du biais dû aux erreurs de classement dans les flux bruts. Lemaître (1988) a effectué une évaluation empirique des méthodes basées sur l'hypothèse de l'indépendance des erreurs de classement et il a montré que ces méthodes peuvent aboutir à une surcorrection du biais de réponse. Gentleman (1988) a examiné les conséquences de certaines hypothèses courantes comme l'absence de biais dans les estimations de stocks et l'indépendance des erreurs de classement, et elle en a conclu que ces hypothèses donnent lieu à des résultats qui ne sont pas confirmés par les données empiriques utilisées dans son étude. Singh et Lemaître (1989) ont essayé d'éviter l'hypothèse de l'indépendance des erreurs de classement en utilisant un modèle de contamination avec paramètre ε pour l'erreur de réponse. Ils ont pris des données provenant d'une autre source, en l'occurrence les données administratives sur les bénéficiaires de l'assurance-chômage, au lieu des données issues des réinterviews, pour estimer le paramètre ε du modèle, qui est analogue aux probabilités d'erreur de réponse. Il convient de noter qu'on n'aurait pas besoin de l'hypothèse de l'indépendance des erreurs s'il était possible de faire des réinterview doubles, c'est-à-dire de réinterviewer les mêmes répondants deux mois de suite, mais on ne l'a pas fait pour des raisons d'ordre opérationnel et pour ne pas alourdir le fardeau de réponse.

Dans le modèle de contamination avec paramètre ε , on a spécifié les erreurs de réponse au moyen d'une distribution de contamination qu'on estime être celle des personnes portées à commettre des erreurs de réponse. On estime cette distribution en faisant l'hypothèse que les estimations de stocks ne sont pas biaisées et que les réponses aux deux moments donnés sont indépendantes uniquement en ce qui concerne les personnes portées à faire des erreurs. Le paramètre ε (approximativement le pourcentage de personnes

link with the true underlying gross flow distribution. Methods for refining the estimate of ϵ and for performing a simulation study for evaluation purposes are currently under investigation; see also Singh and Rao (1990) for further developments.

11.5 Longitudinal Data

The longitudinal data resulting from LFS households remaining in the survey sample for six consecutive months can provide additional information not available in data from a single month. However, most of the estimates and analyses currently use only cross-sectional data. The discovery of a number of possible uses for longitudinal data have led to recent initiatives to begin exploring the largely untapped potential of the longitudinal information.

For example, analysis of response patterns to questions over time can provide new insight into the response mechanism and help to identify problematic questions. This includes potential ambiguities or confusion in concepts or response categories which can be identified by examining responses over time which are highly improbable or logically inconsistent. In this way, the effect of cognitive demands on the respondent to answer difficult or subjective questions, and of demands on the interviewer to interpret and code responses as stated can be assessed. Proxy effects can similarly be studied by examining the longitudinal consistency for proxy versus non-proxy cases.

Longitudinal data can also be used to analyze labour force trends and patterns at the micro-level. Such analyses can yield different results from their aggregate counterparts. The measurement and analysis of gross flows can be extended to longer periods of time, which might, for example, shed light on the degree to which observed flows are concentrated in a small proportion of the population. Borderline job attachment or attachment to the labour force can be better assessed with data over time. This might also indicate important criteria or specific questions affecting changes in labour force status.

Longitudinal data could also be used for editing. This could include the use of on-line edits in a dependent interviewing environment, which might help reduce longitudinal response inconsistencies.

Work is currently focused on the analysis of response patterns over time, and the development of measures of inconsistency for individual questions. Results from this study could serve as valuable input in the planning of any future redesign of the LFS questionnaire and also in improving interviewer instructions. The impact of such inconsistencies on measurements such as estimates for labour force

dans la population qui sont portées à faire des erreurs) sert de lien avec la vraie distribution des flux bruts. Des méthodes visant à raffiner l'estimation de ϵ et à permettre la réalisation d'une étude de simulation à des fins d'évaluation sont actuellement à l'étude; voir aussi Singh et Rao (1990).

11.5 Données longitudinales

Les données longitudinales obtenues du fait que les ménages qui participent à l'EPA font partie de l'échantillon pendant six mois consécutifs peuvent fournir des renseignements de plus que les données produites après un seul mois. Pourtant, la plupart des estimations et des analyses qui sont faites actuellement sont basées uniquement sur les données transversales. Comme il y a plusieurs utilisations possibles pour les données longitudinales, on a commencé ces derniers temps à étudier le potentiel non exploité de ces données.

Par exemple, en analysant l'évolution des réponses dans le temps, on peut acquérir de nouvelles connaissances sur la façon dont les gens répondent et découvrir les questions qui posent des problèmes. En examinant les réponses qui ne sont vraiment pas plausibles ou qui sont en contradiction avec d'autres réponses, on peut mettre en évidence des ambiguïtés éventuelles ou une confusion au niveau des concepts ou des catégories de réponses. On peut alors évaluer l'effet des attentes qu'on a du point de vue cognitif à l'égard des répondants à qui on pose des questions difficiles ou subjectives et à l'égard de l'intervieweur à qui on demande d'interpréter et de coder les réponses. On peut aussi étudier l'effet des réponses par procuration en comparant la cohérence dans le temps de ce genre de réponses par rapport aux réponses données directement par la personne intéressée.

Il est également possible de se servir des données longitudinales pour analyser les tendances de la population active au niveau des microdonnées. Ces analyses peuvent donner des résultats différents de celles réalisées avec les données agrégées. Il est possible de faire porter la mesure et l'analyse des flux bruts sur des périodes plus longues, ce qui pourrait permettre, par exemple, de déterminer dans quelle mesure il y a une concentration des flux observés dans une petite proportion de la population. Les données longitudinales peuvent permettre de mieux évaluer l'attachement à l'emploi ou la participation au marché du travail à caractère marginal, ce qui indiquerait alors quels critères influent sur les changements de situation vis-à-vis de l'activité et quelles questions poser.

Les données longitudinales pourraient aussi servir à la vérification. On pourrait notamment faire des vérifications en direct au cours des interviews avec rétro-information, ce qui réduirait peut-être le nombre de données en contradiction avec celles recueillies auparavant.

Les travaux en cours actuellement font principalement l'analyse de l'évolution des réponses dans le temps et tentent de mettre au point des moyens de mesurer l'incohérence de certaines questions en particulier. Les résultats pourraient s'avérer utiles lorsqu'on planifiera la refonte du questionnaire de l'EPA et qu'on essaiera d'améliorer les instructions à l'intention des intervieweurs. L'effet de telles incohérences sur les mesures comme les estimations relatives aux catégories

status categories and associated gross flows will also be assessed. Plans are in place to investigate the feasibility of on-line longitudinal edits in a CATI environment as part of the research leading up to the next redesign. Imputation methods that incorporate both cross-sectional and longitudinal data are also under investigation.

11.6 Variance Studies

The sample variance and the coefficient of variation are measures of the sampling error associated with survey estimates. Variance estimation procedures for the Labour Force Survey are discussed in sections 5.10 and 9.2. Research is being carried out on alternative variance estimation procedures, particularly for application to quantiles, and to seasonally adjusted data.

Currently, the LFS uses the jackknife method to estimate variances. Although the jackknife is a flexible method, research into its properties has shown that it may not produce appropriate variance estimates for medians and other quantiles. The balanced repeated replication (BRR) method of variance estimation, however, does not have this drawback. Unlike the jackknife, which is based on the omission of one cluster or PSU at a time (see section 8.10), BRR is based on the omission of a complete half-sample at a time. Half-samples are determined by dividing the sample in each stratum into two halves, only one of which is kept for each replication. By choosing many half-samples in a certain pattern, the resulting population estimates can be used to derive a valid variance estimate. A study to compare these variance estimation procedures is in progress.

The LFS is used to obtain both unadjusted and seasonally adjusted estimates of characteristics, such as persons unemployed and in the labour force. Currently, variances are available only for the unadjusted estimates. The derivation of appropriate variances for seasonally adjusted estimates is particularly important since the adjusted estimates are used for analysis and planning and are announced regularly by the news media. Research into variance estimation for seasonally adjusted data is in progress.

11.7 Telephone and CATI Development

Telephone interviewing was first introduced to the Labour Force Survey during the early 1970s. A "warm" telephone interviewing procedure was in place, whereby the initial month's interview was face-to-face, but follow-up interviews in later months, provided the respondent gave his or her consent, were conducted by telephone from the interviewer's home. The telephone procedure was restricted to Self Representing areas. Following successful testing during the post-1981 census redesign, the warm telephoning procedure was extended to remaining areas.

d'activité et aux flux bruts s'y rapportant sera aussi évalué. En prévision du prochain remaniement du plan de sondage, il est question qu'on étudie la faisabilité de vérifications longitudinales en direct dans le contexte de l'interview téléphonique assistée par ordinateur. On examine aussi en ce moment des méthodes d'imputation incorporant à la fois les données transversales et les données longitudinales.

11.6 Études de la variance

La variance de l'échantillon et le coefficient de variation sont les mesures de l'erreur d'échantillonnage dont sont entachées les estimations produites à partir des résultats des enquêtes. Les façons dont on estime la variance pour l'enquête sur la population active sont traitées aux sections 5.10 et 9.2. Des travaux de recherche en cours examinent d'autres méthodes d'estimation de la variance, en particulier pour ce qui a trait aux quantiles et aux données désaisonnalisées.

Actuellement, pour l'EPA, on se sert de la méthode du jackknife pour estimer les variances. Bien que cette méthode soit souple, des travaux de recherche sur ses propriétés ont montré qu'elle ne produit peut-être pas de bonnes estimations en ce qui a trait aux médianes et aux autres quantiles. La répétition compensée est une méthode d'estimation de la variance qui n'a pas cet inconvénient. Contrairement à la méthode du jackknife, selon laquelle il faut omettre une grappe ou une UPE à la fois (voir la section 8.10), la répétition compensée repose sur l'omission d'une moitié entière de l'échantillon à la fois. On obtient la moitié de l'échantillon en divisant en deux l'échantillon dans chaque strate et en n'en gardant qu'une moitié pour chaque répétition. En suivant un certain ordre pour choisir un grand nombre de demi-échantillons, on obtient des estimations de la population dont on peut se servir pour calculer une estimation de la variance qui est valable. Ces deux méthodes d'estimation de la variance font actuellement l'objet d'une comparaison.

On se sert des résultats de l'EPA pour produire à la fois des estimations brutes et des estimations désaisonnalisées relativement à des caractéristiques comme le nombre de personnes en chômage et l'effectif de la population active. À l'heure actuelle, on ne calcule la variance qu'en ce qui concerne les estimations brutes. Or, il est particulièrement important que les variances relatives aux estimations désaisonnalisées soient connues, car ce sont ces estimations qu'on utilise pour les analyses et la planification et que les médias diffusent régulièrement. Des recherches sont en cours à ce sujet.

11.7 Nouvelles façons d'utiliser l'interview téléphonique et l'interview téléphonique assistée par ordinateur

L'interview téléphonique a été introduite dans l'enquête sur la population active au début des années 70. On a choisi une approche selon laquelle on ne procédait pas à froid; c'est-à-dire que, le premier mois, on faisait une interview en personne, pour poursuivre les mois suivants, si le répondant était d'accord, par des interviews téléphoniques que l'intervieweur menait de chez lui. On a limité les interviews téléphoniques aux secteurs autoreprésentatifs. Les essais menés pendant le remaniement du plan de sondage au début des années 80 ayant donné de bons résultats, cette méthode a été étendue aux autres secteurs.

In 1986, a program of research and testing was initiated to investigate new telephone survey methods with a view to reducing survey costs and improving data quality. Telephone studies conducted from 1986-1989 have included: (i) extending telephone interviewing to first-month-in-sample cases, by local interviewers under the area sample design; (ii) conducting the LFS as a central telephone survey with both telephone sampling and telephone collection; (iii) using Computer Assisted Telephone Interviewing (CATI) methods; (iv) using a mixed organization whereby interviewing is done by a combination of local interviewers doing face to face interviewing, and central interviewers telephoning from the regional officer; and (v) studying frame issues.

Findings from these studies are reported by Drew, Dick and Switzer (1989), and are highlighted below. These are followed by comments on the frame and data collection methods likely to be implemented in the post-1991 censal redesign.

Two procedures for telephoning under the area frame were implemented after testing. The first consists of telephone follow-up of first month cases where no contact was made during an initial visit. Interviewers obtain telephone numbers using reverse directories, names on mail boxes, and so forth. This procedure has led to telephoning for 20 percent of first month cases. The second procedure consists of a head office computerized match of addresses for selected dwellings against lists containing published telephone numbers purchased from telephone companies. Where telephone numbers are obtained, the local interviewer conducts first month interviews by telephone, with face to face follow-up of nonresponse. This procedure is scheduled for implementation during 1990 for the apartment frame sample, where it has been shown the procedure leads to 50%-60% of first month cases being interviewed by telephone, with lower nonresponse rates. The procedure is still under evaluation for the remainder of urban areas.

Two tests of the LFS as a centralized telephone survey conducted by interviewers working out of the regional offices were carried out. The tests examined both Random Digit Dialing (RDD) and telephone list sampling methods. Principal findings were that while the costs of the centralized telephone methodology were considerably less than those of the traditional methodology, nonresponse rates were higher - 9% to 12% for the test compared to 4.5% for the ongoing LFS. No significant differences were found between survey estimates for the test treatments, and those for the ongoing LFS adjusted to exclude households without telephones.

A test of Computer Assisted Telephone Interviewing compared the conduct of the LFS as a central telephone survey using CATI methods versus

En 1986, on a lancé un programme de recherche et d'essais visant à étudier de nouvelles méthodes d'interview téléphonique qui permettraient de réduire le coût de l'enquête et d'améliorer la qualité des données. De 1986 à 1989, on y a examiné plusieurs possibilités, dont les suivantes : (i) étendre l'interview téléphonique aux ménages du premier groupe de renouvellement et faire faire celle-ci par des intervieweurs locaux dans les secteurs où l'échantillon est aréolaire; (ii) mener l'EPA en tant qu'enquête téléphonique centralisée où l'échantillonnage et la collecte des données se feraient par téléphone; (iii) utiliser les méthodes d'interview téléphonique assistée par ordinateur; (iv) utiliser une approche mixte, selon laquelle certaines interviews seraient faites sur place par des intervieweurs locaux et certaines seraient faites par téléphone à partir du bureau régional; (v) régler les problèmes liés à la base de sondage.

Drew, Dick et Switzer (1989) ont fait état des résultats de ces études. Nous allons les résumer ci-après et y ajouter des remarques au sujet des méthodes d'établissement de la base de sondage et de collecte des données qui seront probablement adoptées au moment du remaniement du plan de sondage qui suivra le recensement de 1991.

Deux méthodes d'interview téléphonique auprès des ménages de la base aréolaire ont été adoptées à la suite des essais. La première consiste à faire un suivi téléphonique auprès des ménages du premier groupe de renouvellement s'il n'y a pas eu de contact à la première visite. Les intervieweurs obtiennent les numéros de téléphone en consultant, entre autres, les annuaires où les abonnés sont inscrits par ordre de numéro et les noms sur les boîtes aux lettres. Cette méthode a permis de joindre par téléphone 20% des ménages du premier groupe de renouvellement. La seconde consiste à appairer, dans un ordinateur du bureau central, les adresses de certains logements et des listes de numéros de téléphones achetées auprès des compagnies de téléphone. Lorsqu'on trouve le numéro qui correspond à une adresse, l'intervieweur local effectue une interview téléphonique dès le premier mois, et fait un suivi sur place s'il n'obtient pas de réponse. Cette méthode doit être mise en application dans le courant de 1990 auprès des ménages de la base d'appartements, où il a été établi qu'elle permettrait d'interviewer par téléphone 50 à 60% des ménages du premier groupe de renouvellement et qu'elle entraînerait une baisse des taux de non-réponses. L'évaluation de cette méthode pour le reste des régions urbaines n'est pas terminée.

On a réalisé deux essais portant sur la possibilité de faire de l'EPA une enquête téléphonique centralisée dans le cadre de laquelle les intervieweurs mèneraient les interviews du bureau régional. Deux méthodes ont été examinées : la composition aléatoire et l'échantillonnage à partir de listes de numéros de téléphone. Les essais ont montré que la méthode d'interview téléphonique centralisée revient beaucoup moins cher que les méthodes traditionnelles, mais qu'en revanche elle augmente le taux de non-réponse (9 à 12% pendant l'essai, comparativement au taux habituel de 4.5% pour l'EPA). On n'a pas relevé de différences significatives entre les estimations produites à partir des données recueillies dans le cadre des essais et les estimations normales de l'EPA desquelles avaient été exclues les données relatives aux ménages sans téléphone.

Dans un autre essai, on a comparé la tenue de l'EPA en tant qu'enquête téléphonique centralisée selon qu'on avait recours à l'interview téléphonique assistée par ordinateur ou

using paper and pencil methods. While there were no cost advantages to CATI over the centralized methodology without CATI, CATI did result in improved data quality. Under CATI there were fewer edit failures, fewer branching errors, and more complete enumeration of persons within households.

A study of the organization of data collection showed that under a mixed organization with central telephone interviewers and local interviewers to handle non-telephone households and face to face follow-up of telephone non-respondents, non-response rates close to those of the traditional methodology could be obtained, while achieving a 14% reduction in data collection costs.

In terms of frame issues, analysis of telephone coverage in Canada revealed that 2% or fewer households do not have telephones. Those without telephones have markedly different labour force characteristics, with higher unemployment rates and lower participation rates. Given the accuracy required for national LFS estimates, telephone frames do not provide adequate coverage. Studies are being conducted into dual frame methods combining telephone and area frames, and in urban areas, into creation of an Address Register – a comprehensive list frame containing residential addresses and published telephone numbers.

Although some new telephone methods have already been introduced as a result of the testing programme, the major benefit has been to learn and build a composite view of the future design and data collection methodology to be implemented during the post-1991 censal redesign.

The methodology envisioned is a mixed one with respect to frame, mode of collection, organization, and possibly technology. The frame is likely to be a dual (area-telephone) frame in rural areas, and an Address Register in urban areas; mode and organization collection will likely feature face to face interviewing of non-telephone households and telephone non-respondents by local interviewers and telephone interviewers working out of the regional offices; and technology is likely to be CATI for the centralized telephone interviewers, and paper and pencil or Computer Assisted Personal Interviewing (CAPI) for the local interviewers.

These components have yet to be tested together, but such integrated testing is planned in the research phase of the next redesign.

11.8 Address Register

The Address Register is a machine-readable list of residential addresses in urban areas of Canada. It is being created by merging and unduplication of address information on administrative records such as telephone and electric utility company billing files, and municipal assessment rolls. It is being developed

que l'interview était menée normalement avec papier et crayon. L'interview téléphonique assistée par ordinateur ne présentait aucun avantage du point de vue des coûts, mais elle améliorait la qualité des données. On relevait moins d'erreurs de vérification et de branchement et le dénombrement des membres des ménages était plus complet.

Il est ressorti d'une étude de l'organisation de la collecte des données que, lorsqu'on a une organisation mixte avec centralisation des interviews téléphoniques et décentralisation des interviews pour les ménages sans téléphone et pour les interviews sur place auprès des personnes n'ayant pas répondu à l'interview téléphonique, on obtient des taux de non-réponse voisins de ceux qu'on a avec la méthode traditionnelle tout en réduisant de 14% le coût de la collecte des données.

Pour ce qui est de la question de la base de sondage, une étude du nombre d'abonnés au Canada a montré que 2% des ménages ou moins étaient sans téléphone. Ces personnes ont des caractéristiques relatives à l'activité nettement différentes de celles du reste de la population : taux de chômage plus élevé et taux d'activité plus faible. Compte tenu des exigences de l'enquête en matière de fiabilité des estimations nationales, les bases de numéros de téléphone ne sont pas suffisamment exhaustives. On est en train d'étudier la possibilité de constituer une double base de sondage combinant les bases de numéros de téléphone et les bases aréolaires et de créer, pour les régions urbaines, un registre d'adresses, c'est-à-dire une liste complète des adresses domiciliaires et des numéros de téléphone publiés.

Le programme d'essais a déjà abouti à l'adoption de nouvelles façons d'utiliser l'interview téléphonique, mais son principal intérêt a été de permettre d'envisager les multiples possibilités ayant trait au plan de sondage et aux méthodes de collecte de données qui seront mis en oeuvre à l'occasion du remaniement faisant suite au recensement de 1991.

La méthodologie envisagée est une méthodologie mixte quant à la base de sondage, au mode de collecte, à l'organisation et peut-être à la technologie. Il s'agira probablement : d'une base de sondage qui sera une base double (base aréolaire et base de numéros de téléphone) dans les régions rurales et un registre d'adresse dans les régions urbaines; d'un mode de collecte qui comportera, d'une part, des interviews en personnes par des intervieweurs locaux auprès des ménages sans téléphone et des personnes qui n'auront pas répondu à l'interview téléphonique et, d'autre part, des interviews téléphoniques effectuées à partir des bureaux régionaux; d'une organisation selon laquelle les intervieweurs opérant des bureaux régionaux auront recours à l'interview téléphonique assistée par ordinateur et les intervieweurs locaux, à leur crayon et au papier ou à l'interview sur place assistée par ordinateur.

Toutes ces composantes n'ont pas encore été testées, mais des essais intégrés sont prévus pour l'étape du prochain remaniement qui sera consacrée à la recherche.

11.8 Registre d'adresses

Le registre d'adresses est une liste ordiolingue d'adresses domiciliaires pour les régions urbaines canadiennes. Pour le constituer, on est en train de fusionner, en évitant tout dédoublement, les renseignements sur les adresses qui se trouvent dans des dossiers administratifs comme les fichiers de facturation des compagnies de

initially for use in the 1991 Census as an independent check on the address lists created in the field by Census enumerators. Studies during the research and testing leading up to the 1991 Census demonstrated that use of the Address Register in this fashion could significantly reduce dwelling undercoverage in the Census (Drew, Royce, and van Baaren 1989).

As part of its use during the 1991 Census, the Address Register will be updated to reflect additions and deletions found during the Census. Following 1991, it is planned to update the Address Register on an ongoing basis, through use of current administrative records, and to adopt the Address Register as a frame for household surveys in urban areas. In addition to being a highly complete list frame, the Address Register will contain telephone numbers in cases where the telephone numbers are published.

The presence of telephone numbers for most units on the frame will permit the majority of households to be interviewed exclusively by telephone, while permitting face-to-face follow-up of telephone non-respondents.

Research is required into efficient sample designs based on the Address Register frame. Issues to be examined include gains from having telephone and non-telephone strata, the extent of clustering of the sample for each stratum, the desirability of stratification of blocks or even dwellings on the basis of Census characteristics, population coverage of the Address Register and whether sampling strategies will be needed to augment its coverage, and sampling strategies for dealing with frame changes.

téléphone et d'électricité et les rôles d'évaluation municipaux. On veut d'abord s'en servir en prévision du recensement de 1991 pour contrôler les listes d'adresses dressées par les recenseurs sur le terrain. Les études menées à cet égard ont montré que l'utilisation d'un registre d'adresses à cette fin pouvait réduire considérablement le sous-dénombrement du recensement (Drew, Royce et van Baaren, 1989).

Le registre d'adresses sera mis à jour en cours d'utilisation au recensement de 1991. Il est prévu que la mise à jour se poursuivra par la suite de façon continue au moyen des dossiers administratifs courants et que le registre d'adresses servira de base de sondage pour les enquêtes-ménages réalisées dans les secteurs urbains. Outre le fait que le registre d'adresses constituera une base de sondage très complète, il contiendra des numéros de téléphone qui normalement ne sont pas publiés.

Comme on disposera d'un numéro de téléphone pour la plupart des unités de la base de sondage, on pourra mener exclusivement des interviews téléphoniques auprès de la majeure partie des ménages et ne faire des interviews sur place qu'auprès des non-répondants.

Il faut faire de la recherche pour mettre à point des plans de sondage efficaces ayant recours à un registre d'adresse comme base de sondage. Il s'agit de déterminer notamment : les gains qui découlent du fait de constituer des strates à interviews téléphoniques et des strates à interviews autres que téléphoniques; le nombre de grappes à former dans l'échantillon de chaque strate; l'avantage de stratifier les îlots ou même les logements en fonction des caractéristiques de recensement; l'exhaustivité du registre d'adresse; et enfin la nécessité éventuelle de mettre au point des techniques d'échantillonnage qui permettraient de rendre le registre plus complet et de réagir aux changements dans la base de sondage.

Appendix A

Numbering System

1. Introduction

The numbering system assigns a unique 17-digit number to each individual in the sample and contains all the design information required for estimation and tabulation. Since each individual is uniquely identified, it also permits the linking of data for individuals across months for the purpose of longitudinal analysis. In brief, the 17-digit numerical identification can be broken down into seven fields, which are described below.

Field No.	Position	Name	Identifies
N° de zone		Nom	Description
1	1-5	PSU – UPE	Province, Economic Region, Stratum and Primary Sampling Unit. – Province, région économique, strate et unité primaire d'échantillonnage
2	6-7	Group – Groupe	Second Stage Sampling Unit in NSR Rural Areas. – Unité du deuxième degré d'échantillonnage dans les régions rurales des secteurs NAR.
3	8-10	Cluster – Grappe	Second Stage Sampling Unit in NSR Urban and in SR areas. – Unité du deuxième degré d'échantillonnage dans les régions urbaines des secteurs NAR et les secteurs AR.
4	11	Rotation – N° de renouv.	Number governing month in which sample rotation occurs. – Chiffre qui indique quel mois le renouvellement de l'échantillon a lieu.
5	12-14	List – Liste	Dwelling List Number. – Numéro de liste du logement.
6	15	Multiple	Additional dwellings within selected units found at interview time. – Logements supplémentaires dans les unités échantillonnées dont on prend connaissance au moment de l'interview.
7	16-17	Page-line – Page/ligne	Person as listed on the Household Record Docket. – Personne telle qu'identifiée dans le dossier du ménage.

The first digit of the PSU consists of the province as follows:

- 0 Newfoundland
- 1 Prince Edward Island
- 2 Nova Scotia
- 3 New Brunswick
- 4 Quebec
- 5 Ontario
- 6 Manitoba
- 7 Saskatchewan
- 8 Alberta
- 9 British Columbia

Although the LFS does not sample in the Yukon or Northwest Territories, for the purposes of testing or for special surveys using the LFS system, the Yukon is usually assigned to Province 9 and the Northwest Territories to Province 8. This corresponds to the regional office designated to cover each area.

Annexe A

Système de numérotation

1. Introduction

Le système de numérotation attribue un numéro à dix-sept chiffres à chaque personne qui fait partie de l'échantillon, et il contient tous les renseignements relatifs au plan de sondage nécessaires à l'estimation et à la totalisation des données. Comme chaque personne a un numéro qui lui est propre, ce système permet de faire le couplage des données d'un mois à un autre à des fins d'analyse longitudinale. Le numéro d'identification se compose de sept zones dont la description est la suivante.

Le premier chiffre de la zone réservée à l'UPE désigne la province, et les codes employés sont les suivants :

- 0 Terre-Neuve
- 1 Île-du-Prince-Édouard
- 2 Nouvelle-Ecosse
- 3 Nouveau-Brunswick
- 4 Quebec
- 5 Ontario
- 6 Manitoba
- 7 Saskatchewan
- 8 Alberta
- 9 Colombie-Britannique

On ne prélève pas d'échantillon pour l'EPA au Yukon ni dans les Territoires du Nord-Ouest, mais pour les essais et les enquêtes spéciales menées au moyen de l'EPA, le Yukon est généralement associé à la province 9 et les Territoires du Nord-Ouest, à la province 8, car ce sont les bureaux régionaux situés dans ces provinces qui s'occupent respectivement de ces territoires.

The second digit of the PSU is the Economic Region. One anomaly exists in this identification. The numbering system was originally designed to have only one digit for the ER, allowing for ten ERs per province. In the current design there are 14 ERs in Quebec and 11 in Ontario, necessitating the collapse of some ERs.

The weighting system has been converted to allow a three-digit Economic Region (including province) to be used in the final estimation and tabulation procedures. In order to assign ERs for those PSUs where the numbering system has been collapsed a table has been built to convert PSUs to ERs. Creating this table was more expedient than changing the numbering system due to the impact on all systems and databases in the LFS. Those PSUs not found in this table have ER values that are generally written as three-digit numbers, the first two digits from the PSU and the third digit zero.

Le deuxième chiffre de la zone UPE représente la région économique. Il y a une anomalie à cet égard. À l'origine, on n'avait prévu qu'un chiffre pour les régions économiques parce qu'on pensait qu'il n'y en aurait que 10 par province. Dans le plan actuel, il y a quatorze RE au Québec et onze en Ontario, de sorte qu'on est obligé de grouper certaines d'entre elles.

Le système de pondération a été modifié de manière à ce qu'on puisse attribuer un code à trois chiffres aux régions économiques (province incluse) au moment de la production des dernières estimations et totalisations. Afin qu'il soit possible d'attribuer un code de région économique aux UPE qui ont été regroupées pour les besoins du système de numérotation, un tableau de conversion a été mis au point (voir en annexe). Il était plus facile de procéder ainsi que de modifier le système de numérotation, car cela aurait eu des conséquences sur tous les systèmes informatiques et bases de données de l'EPA. Le code des UPE qui ne se trouvent pas dans le tableau est généralement un numéro à trois chiffres, dont les deux premiers identifient l'UPE et le troisième est un zéro.

PSU Ranges with Non-Standard Economic Regions

PSU Range		ER
Tranche de UPE		RE
41000	41019	411
41020	41059	412
41100	41199	412
41200	41299	411
46000	46039	461
46040	46059	463
46060	46099	464
46100	46199	462
46200	46219	461
46220	46239	462

Tranches de UPE pour les régions économiques avec code à trois chiffres

PSU Range		ER
Tranche de UPE		RE
46240	46249	463
46250	46299	464
46300	46699	461
46700	46799	463
46800	46809	461
46810	46899	464
59000	59039	591
59040	59059	592
59100	59699	591
59700	59899	592

2. Basic LFS Strata

The detailed significance of the first three fields varies according to the basic type of design stratum involved. There are four basic types: NSR, SR non-apartment, SR apartment and Special Areas. These area types can be identified by the sixth and seventh digits as shown in the table below.

Note that the 50s are not mentioned in the table. This range was reserved for special categories of non-LFS units that may be added to the database from time to time. For instance, the Family Expenditure Survey regularly conducts interviews in both Whitehorse, Y.T., and Yellowknife, N.W.T. and has been assigned group numbers in this range.

It is also worth noting that the third digit identifies NSR, SR and Special Areas but does not distinguish SR apartment from non-apartment strata.

2. Strates de base de l'EPA

La signification des trois premières zones varie selon le genre de strate dont il s'agit. Fondamentalement, il y a quatre genres de strates : les secteurs NAR, les secteurs AR sans la base d'appartements, les secteurs AR avec base d'appartements et les secteurs spéciaux. Le genre de strate est indiqué par le sixième et le septième chiffres, comme on peut le voir dans le tableau qui suit.

On remarquera que la cinquantaine est absente du tableau. Elle a été réservée à des catégories spéciales de groupes non sélectionnés pour l'EPA qui peuvent être ajoutés à la base de données à l'occasion. Ainsi, pour l'enquête sur les dépenses des familles, il se fait régulièrement des interviews à Whitehorse (Yukon) et à Yellowknife (T.N.-O.) et on attribue aux groupes sélectionnés un code entre 50 et 60.

Il convient de signaler également que le troisième chiffre identifie les secteurs NAR, AR et spéciaux mais qu'il ne fait pas la distinction entre les strates AR formées sans la base d'appartements et celles formées à partir de la base d'appartements.

Type of Area	Sixth-Seventh Digits	Third Digit
Genre de secteur	Sixième et septième chiffres	Troisième chiffre
NSR – NAR	11-49	0
SR – AR	61-84	1-8
Apartment – Base d'app.	91-96	1-8
Special Area – Spécial	01-10	9

In the following sections the detailed numbering system for each area type is described.

Le système de numérotation pour chaque genre de secteur est décrit en détail dans les sections qui suivent.

3. NSR Areas

As in previous designs the stratification of NSR areas is designated by ranges of PSUs. The last two digits of the PSU field are interpreted as follows:

Stratum	Range
Strate	Tranche
1	01-19
2	21-39
3	41-59
4	61-79
5	81-99

A new PSU category has a value of 99 in the last two digits. Seven PSUs in the LFS have a special design as follows:

3. Mode de numérotation des UNAR

Comme dans l'ancien système, la stratification des secteurs NAR est représentée par des tranches de numéros dans la zone réservée aux UPE. Voici comment interpréter les chiffres dans ces tranches, en avant-dernière et dernière position de la zone UPE :

Il y a une nouvelle catégorie d'UPE dont les deux derniers chiffres sont 99. Sept UPE de l'EPA ont un plan de sondage spécial. Ce sont :

PSU	Design
UPE	Plan
25099	Random Group/Six selections – Groupe aléatoire/six unités
53099	Random Group/Six selections – Groupe aléatoire/six unités
57099	Buffer/Random Group/Two selections – Tampon/groupe aléatoire/ deux unités
73099	Buffer/Random Group/Two selections – Tampon/groupe aléatoire/ deux unités
83099	Buffer/Random Group/Two selections – Tampon/groupe aléatoire/ deux unités
86099	Buffer/Random Group/Two selections – Tampon/groupe aléatoire/ deux unités
98099	Random Group/Six selections – Groupe aléatoire/six unités

The buffer areas were set up to provide the capacity of updating NSR areas just outside of major cities where no rural census subdivision was included in the CMA boundaries. Two EAs are selected within the buffer area by the random group method. The buffer around Windsor (57099) allows a more equitable comparison between the current design estimates and the old design estimates for this CMA. The CMA associated with these buffers start with a PSU number equal to the buffer PSU number plus one.

On a créé des zones tampons pour permettre de faire la mise à jour de l'échantillon dans les secteurs NAR juste à l'extérieur des grandes villes dans les limites desquelles il n'y avait pas de subdivisions de recensement rurales. Deux SD ont été choisis dans ces zones tampons par la méthode des groupes aléatoires. La zone tampon autour de Windsor (57099) permet de faire des comparaisons plus justes entre les estimations relatives à cette RMR selon le plan de sondage actuel et les estimations selon l'ancien plan de sondage. Le code des RMR associées à ces zones tampons commence par un code d'UPE qui est le code d'UPE de la zone tampon plus 1.

The other PSUs in this category are in Economic Regions that have an insufficient NSR population for normal PSU formation. Six EAs are selected with the random group method. All seven PSUs in this new category are strictly rural.

The group number identifies the area sub type. Groups 11-19 are urban groups and 21- 49 are rural groups. In urban areas the clusters generally correspond to city blocks and are numbered from 001 to the number of clusters formed in that town or city. In rural areas most clusters will be 000 corresponding to the EA design in this area (i.e., no clustering). If the EA had to be split after selection of the group to provide 6 rotations per PSU, then the two clusters so formed would be numbered 001 and 002 and both would be selected.

In the PSUs ending in 99, the group refers to the random group to which the cluster was assigned; the cluster identifies the selected EA.

4. SR Areas

The first three digits of the PSU will generally identify the city involved. This is the SRU number, often called the SRU. For example, SRU 011 is St. John's, NFLD. It is the first or largest city in Economic Region 010. All identifications starting with 011 will be located within the CMA of St. John's. The largest cities require more than one SRU number, as follows:

CMA	SRU Numbers
RMR	Code d'UAR
Montreal	461, 462
Toronto	531, 536
Winnipeg	611, 661, 671, 672
Edmonton	861, 862
Vancouver	951, 955

Another exception exists in SRU 468. The numbering system allows eight different SRU codes within the two-digit ER 46. There are, in fact, four ERs starting with 46: 461;462;463;464. The SRU sample in these four ERs consists of eight cities, but Montreal requires two SRU codes as shown above. Hence, the smallest two cities have to share the SRU number as follows:

PSU	City
UPE	Ville
46801-46802	Valleyfield
46811	Joliette

Please note that the third digit equal to 9 is already reserved for the Special Areas.

Les autres UPE sont dans des régions économiques dont les secteurs NAR ont une population insuffisante pour qu'on puisse y former des UPE normalement. Six SD sont choisis par la méthode des groupes aléatoires. Les sept UPE de cette nouvelle catégorie sont entièrement rurales.

Le code de groupe identifie le genre de sous-secteur. Les groupes 11 à 19 sont des groupes urbains et les groupes 21 à 49, des groupes ruraux. Dans les régions urbaines, les grappes sont généralement des îlots, et elles sont numérotées à partir de 001 jusqu'à concurrence du nombre de grappes formées dans la ville en question. Dans les régions rurales, le code de grappe est la plupart du temps 000, compte tenu du plan de sondage des SD (pas de grappe). Si, après la sélection du groupe, il a fallu diviser le SD pour obtenir six groupes de renouvellement par UPE, les deux grappes ainsi formées portent les codes 001 et 002 et elles sont choisies toutes les deux.

Pour les UPE dont le code finit par 99, le code de groupe désigne le groupe aléatoire auquel la grappe a été attribuée et le code de grappe, le SD choisi.

4. Mode de numérotation des UAR

Les trois premiers chiffres de la zone UPE identifient généralement la ville. Ils forment le code de l'UAR. Par exemple, le code d'UAR 011 désigne St. John's (Terre-Neuve). C'est la première ou la plus grande ville de la région économique 010. Toutes les unités d'échantillonnage dont le code commence par 011 sont situées dans la RMR de St. John's. Les grandes villes représentées par plus d'un code d'UAR sont les suivantes.

L'UAR 468 constitue une exception. Le système de numérotation prévoit huit codes d'UAR différents pour le code de RE 46. Dans les faits, il y a quatre RE dont le numéro commence par 46, les RE 461, 462, 463 et 464, et huit UAR (villes) dans ces quatre RE; cependant, comme nous l'avons vu plus haut, il faut deux codes d'UAR pour Montréal. Les deux plus petites villes doivent donc partager le code d'UAR 468.

Il faut noter que le chiffre 9 en troisième position est déjà réservé aux secteurs spéciaux.

The fourth digit of the PSU number identifies a contiguous geographic stratum within certain SRUs. The larger cities can have as many as 19 of these "super strata", hence the need for more than one SRU number. The fifth digit identifies the sub-unit (also called PSU) that was formed within these super strata. The SRU design limits the fifth digits to 1-9 since 0 is reserved for the Apartment frame. The number of sub-units in a super stratum is typically between three and five.

Small cities do not have a large enough sample for this level of stratification. All but one have the fourth digit equal to zero. The exception is another anomaly. Fredericton is too small to have super strata and yet it was designed with eleven PSUs numbered consecutively from 34101 to 34111. Thus it looks like it has two super strata, which is not true, and PSU 34110 looks like an Apartment frame PSU, which it is not.

As before the group identifies a random subset of the clusters in a PSU and only one cluster is selected per random group. In the outskirts of larger cities a cluster can be an entire EA; elsewhere, it is typically a city block.

5. Apartments

As in SRUs, the first four digits of the PSU identify the location of the Apartment building in terms of the super stratum. The fifth digit is always zero. (PSU 34110 is the only exception to the reverse rule that Self-Representing PSUs ending in zero are part of the Apartment design.)

The PSU number in this design is only a location identifier since the sampling scheme does not use the PSU number in the stratification process. Generally, all apartment buildings in a city are listed sequentially, with cluster numbers 001 to the number of buildings in total. The group numbers alternate between 91 and 92 for each successive cluster selection.

In some cities there are enough clusters to assign size strata in this frame. The group number identifies the size stratum as follows:

Dans certaines UAR, le quatrième chiffre de la zone UPE désigne une strate constituée de secteurs de recensement contigus. Les grandes villes peuvent avoir jusqu'à 19 de ces "superstrates". Elles ont par conséquent besoin de plus d'un code d'UAR. Le cinquième chiffre identifie les sous-unités (aussi appelées UPE) formées à l'intérieur de ces superstrates. Selon le système de numérotation, le cinquième chiffre peut varier de 1 à 9 seulement, le chiffre 0 étant réservé à la base d'appartements. En général, il y a entre trois et cinq sous-unités dans une superstrate.

L'échantillon prélevé dans les petites villes est trop petit pour qu'on puisse le soumettre à ce niveau de stratification. Toutes sauf une ont le chiffre 0 en quatrième position. L'exception constitue encore une fois une anomalie. Fredericton est trop petite pour avoir une superstrate, et pourtant on y a créé 11 UPE ayant pour code les numéros consécutifs allant de 34101 à 34111. On a donc l'impression qu'elle a deux superstrates, ce qui n'est pas le cas, et on pourrait croire que l'UPE 34110 a été prélevée dans une base d'appartements, ce qui n'est pas le cas non plus.

Comme dans les UNAR, le code de groupe identifie un sous-ensemble aléatoire de grappes dans une UPE, et une grappe seulement est choisie par groupe aléatoire. Dans la périphérie des grandes villes, la grappe peut être un secteur de dénombrement complet; sinon, c'est en général un îlot.

5. Mode de numérotation des UAR de la base d'appartements

Comme dans le cas des UAR, les quatre premiers chiffres de la zone UPE indiquent l'emplacement de l'immeuble d'appartements du point de vue de la superstrate. Le cinquième chiffre est toujours un zéro. (Nota : l'UPE 34110 est la seule exception à la règle selon laquelle les UPE des UAR dont le code finit par zéro ont été formées dans la base d'appartements.)

Dans ce mode de numérotation, les chiffres dans la zone UPE permettent uniquement d'identifier l'emplacement, car la méthode d'échantillonnage n'utilise pas le code d'UPE aux fins de la stratification. En général, on inscrit les immeubles d'appartements d'une ville sur la liste dans l'ordre, l'un après l'autre, et on leur attribue un code de grappe en commençant par 001 et en allant jusqu'au nombre total d'immeubles dans la liste. Le code de groupe alterne entre 91 et 92 à chaque grappe échantillonnée.

Dans certaines villes, il y a suffisamment de grappes dans cette base pour qu'on puisse former des strates selon un critère de taille. Ces strates sont représentées par les codes de groupe suivants.

Group Numbers	Standard Size Stratum	Toronto Size Stratum
Codes de groupe	Taille standard des strates	Taille des strates à Toronto
91-92	30-99 units - unités	30-99 units - unités
93-94	100 + units - unités et plus	100-199 units - unités
95-96	-	200 + units - unités et plus

In Toronto and Montreal there are also geographic strata identified by sets of super strata. The use of super stratum for the PSU field implies that "PSU + Cluster" is no longer a unique identifier, as it is in the SRU. Group numbers are required to identify the size stratum. The advantage appears when the clusters are assigned to interviewers at the super stratum level.

6. Special Areas

The PSU number can only identify the ER where the Special Area unit is located. In most areas the frame is not stratified at the ER level. The second digit is almost always zero. The exception is in the design of the Alberta-remote, where the second digit can have two possible values of 7 and 8 corresponding to the two remote strata in that province.

The ER number for each Special Area unit will be identified by digits 4 and 5 of the PSU. This, combined with the first digit (Province), will yield the full three-digit Economic Region.

The two major area subtypes are denoted by the group number: 01 defines the Military/Institutional frame and 03 defines the Remote Area frame. Additionally, the special Quebec Urban Remote frame uses group numbers 04 to 10, one for each urban area in the frame.

The cluster number determines the EA selection within a Special Area stratum. The clusters are assigned sequentially on a list of the EAs sorted by descending order of size. If one of the selected EAs is actually clustered, the cluster number is revised to a new value outside the range of the main list of EAs. Usually the first clustered selection is numbered from 101 to 106, for example, and the second may have 201-202. (Note: Very few Special Area EAs are actually clustered; all are remote.)

In the Quebec Urban Remote frame each of the two selected urban areas is divided into six clusters, all of which are selected and numbered sequentially from 001-006.

7. Rotation Number

The rotation number is also known as the rotation group or just the rotation. It governs the month of introduction of the selected dwellings within the cluster.

A Toronto et à Montréal, il y a aussi des strates géographiques représentées par des tranches de chiffres à l'intérieur du code des superstrates. L'utilisation du code de superstrate dans la zone d'UPE signifie que le code d'UPE et le code de grappe ne constituent plus, ensemble, un identificateur unique, comme dans le cas des UAR. On a besoin du code de groupe pour identifier la strate formée selon le critère de taille. L'avantage de ce mode de numérotation ressort lorsque les grappes sont attribuées aux intervieweurs au niveau de la superstrate.

6. Mode de numérotation des secteurs spéciaux

Les chiffres inscrits dans la zone UPE indiquent uniquement la RE où le secteur spécial échantillonné est située. Dans la plupart des secteurs, la base n'est pas stratifiée au niveau de la RE. Le deuxième chiffre est presque toujours zéro. La seule exception est la base des régions éloignées de l'Alberta, où le deuxième chiffre, qui représente les deux strates de régions éloignées de cette province, peut avoir deux valeurs possibles, 7 ou 8.

Les chiffres en quatrième et en cinquième position de la zone d'UPE identifient la RE de chaque unité des secteurs spéciaux. Combinés au premier chiffre (code de la province), ils constituent le code à trois chiffres complet de la région économique.

Le code de groupe identifie les deux principaux types de sous-secteurs : 01 désigne la base militaire/institutionnelle et 03, la base des régions éloignées. Il y a en outre la base spéciale des régions éloignées du Québec, dont chaque région urbaine porte un code de groupe allant de 04 à 10.

Le code de grappe désigne le SD choisi dans la strate des secteurs spéciaux. Ce code est un numéro séquentiel attribué aux SD classés par ordre décroissant de taille. Si un des SD choisis est subdivisé en grappes, on attribue à celles-ci un code de grappe dont la valeur ne figure pas dans la principale liste de SD. En général, on attribue au premier SD subdivisé en grappes les codes 101 à 106 et au deuxième, les codes 201 et 202. (Il convient de noter que très peu de SD dans les secteurs spéciaux sont subdivisés en grappes. Il s'agit dans tous les cas de régions éloignées.)

Dans la base des régions éloignées du Québec, les deux régions urbaines choisies sont subdivisées en six grappes qui sont toutes échantillonnées et qui sont numérotées dans l'ordre de 001 à 006.

7. Numéro de renouvellement

Le numéro de renouvellement, attribué aux grappes, détermine le mois où les logements choisis dans les grappes sont introduits dans l'échantillon. Les mois auxquels correspondent les numéros de renouvellement sont indiqués ci-après.

The correspondence is shown below:

Rotation Number	Months of Introduction
Numéro de renouvellement	Mois d'introduction
1	January, July – Janvier, juillet
2	February, August – Février, août
3	March, September – Mars, septembre
4	April, October – Avril, octobre
5	May, November – Mai, novembre
6	June, December – Juin, décembre

For each month of the LFS, one of the six rotations thus represents all the dwellings that have just been introduced for this month. This is variously called the birth rotation for the month, the birth month for the dwelling and the birth dwellings for the month.

Some of the dwelling lists represented by PSU-Group-Cluster may be selected in two different rotations over the life of a PSU. Typically this occurs in NSR Rural Groups. The identifications are different and yet they physically share the same list of dwellings on the LFS data base. They are called linked or shared clusters. The separate identifications allows the double sampling of the list, although care has to be taken to avoid overlapping the dwelling selections.

The cluster identification specifically refers to the third field of the numbering system. In general, the entire set of PSU-Group-Cluster-Rotation is also called the cluster since this set of numbers is used as the key field for the LFS data base.

8. Cluster Lists

In the normal process the land area (or apartment building or institution) represented by the PSU-Group-Cluster is investigated in the field and every dwelling found in the area is listed with a unique civic address or an address description. The listing number identifies the sequence of dwellings on this list. Each listing number represents a separate dwelling which may or may not contain a household to be interviewed.

Occasionally the person listing the dwelling does not realize that there more than one dwelling located at that address. Upon interviewing in the dwelling all such extra dwellings are listed on a separate line for data capture but with the same listing number. The multiple number allows up to nine such dwellings to be captured for each original listing.

Chaque mois où l'EPA a lieu, un des six numéros de renouvellement désigne tous les logements introduits dans l'échantillon ce mois-là. On parle dans ce contexte du numéro de renouvellement des logements introduits ce mois-là, du mois de renouvellement des logements et des logements introduits dans l'échantillon ce mois-là.

Il arrive que certaines listes de logements portant un code d'UPE, de groupe et de grappe donne fassent partie de deux groupes de renouvellement différents pendant la période active d'une UPE. Cela se produit généralement dans les groupes ruraux des secteurs NAR: leur numéro d'identification est différent et pourtant ils figurent sur la même liste de logements dans la base de données de l'EPA. On les appelle des grappes enchainées. Le fait de leur donner un numéro d'identification différent permet d'échantillonner la liste deux fois, quoiqu'il faille faire attention de ne pas prélever les mêmes logements deux fois.

Le code de grappe est la troisième zone du système de numérotation. En général, on appelle aussi le numéro forme par le code d'UPE, le code de groupe, le code de grappe et le numéro de renouvellement le code de grappe, car il s'agit de la zone clé en ce qui concerne la base de données de l'EPA.

8. Listes de grappes

Selon le processus normal, toute superficie (ou immeuble d'appartements ou établissement) représentée par un code d'UPE, de groupe et de grappe est visitée par le personnel sur le terrain qui dresse la liste des logements qui s'y trouvent en indiquant pour chacun soit son adresse de voirie, soit une description de l'endroit où il est situé. Le numéro de liste suit l'ordre dans lequel les logements sont inscrits sur la liste. Chaque numéro de liste correspond à un logement distinct qui peut contenir ou non un ménage à interviewer.

Il arrive que la personne chargée de dresser la liste de logements ne se rende pas compte qu'il y a plus d'un logement à une adresse donnée. Lorsqu'on interview le ménage, on ajoute les autres logements sur une ligne distincte pour la saisie des données mais on leur attribue le même numéro de liste. Le code attribué au logement additionnel découvert dans un immeuble à logements multiples permet de saisir les données se rapportant à un maximum de neuf logements par logement relevé au départ.

Sampling of the cluster lists generates a set of dwellings represented by identification of PSU-Group-Cluster-Rotation-Listing-Multiple. This identification is then converted to a separate but also unique numbering system listing the Regional Office and Docket Number (RO-Docket). These docket numbers are sequentially assigned as the sample is added to the Response file that controls the flow of documents from this point onward. A check digit is added after the sequential number to provide an eight-digit reference item.

9. Individuals

One can use either the PSU-Group-Cluster-Rotation-Listing-Multiple or the RO- Docket number to proceed from the household identification to the individual. Within the household, all individuals are listed on the Household Record Docket on one or more pages, with each individual on a separate line of the docket. Thus the identification refers to these individuals as page-line entries.

Lorsqu'on échantillonne les listes de grappes, on obtient un ensemble de logements représentés chacun par un numéro composé des codes d'UPE, de groupe, de grappe, de renouvellement, de liste et de logement dans un immeuble à logements multiples (UPE-Groupe-Grappe-Renouvellement-Liste-Multiple). Ce numéro d'identification est converti en un autre numéro également unique, le numéro de bureau régional et de dossier (BR-Dossier). Ce numéro d'ordre est attribué aux unités de l'échantillon au fur et à mesure qu'on les ajoute au fichier des réponses qui contrôle le flux des documents à partir de ce point. Un chiffre de contrôle est ajouté à ce numéro d'ordre, ce qui donne un numéro de référence à huit chiffres.

9. Individus

Il est possible d'utiliser les codes des "UPE-Groupe-Grappe-Renouvellement-Liste-Multiple" ou les codes des "BR-Dossier" afin d'identifier les individus d'un ménage. Tous les individus d'un même ménage sont énumérés dans le dossier du ménage sur une ou plusieurs pages en inscrivant un individu par ligne. Ainsi, l'information relative à chaque individu peut être référée par un numéro de page et de ligne.

Appendix B

Sample Size Stabilization

1. Introduction

The stabilization program was instituted in 1977 to limit the growth of the LFS sample. The number of newly sampled dwellings (births) selected each month was limited to a previously set value called the base figure. In the previous design each major category of the sample within each province was given a separate base figure. The categories were

Self-Representing Non-Apartment (SR Regular)

Self-Representing Apartment (SR Apt)

Non-Self-Representing Urban (NSR-Urban)

Non-Self-Representing Rural (NSR-Rural) and

Special Areas (SPE).

Not all areas were stabilized. All Special Areas were left unstabilized as were some of the smaller SR areas. The definitions of the areas were hard coded into the software that performed the sampling of dwellings from LFS clusters. In order to alter the definition of stabilization areas a program modification proposal was necessary. This ensured that all components of the system received the same changes.

The actual dropping of dwellings was performed according to the following algorithm. First, the off-rotation clusters and the clusters with sub-weights due to large growth were identified. Off-rotation clusters are those introduced in a month other than that dictated by the rotation number. Such clusters occur mostly in the open-ended Apartment frame in order to sample dwellings as soon as newly added clusters become selected.

Clusters with sub-weights were counted in the sample size for the stabilization area but no dwelling selections within these clusters were eligible for dropping and none of these clusters actually received the stabilization weight. The stabilization weight applied to the remaining clusters was calculated using

where:

W is the weight

N is the original count of selected dwellings

B is the base figure

C is the count of selected dwellings in sub-weighted clusters and

Annexe B

Stabilisation de la taille de l'échantillon

1. Introduction

Le programme de stabilisation a été institué en 1977 pour limiter la croissance de l'échantillon. Le nombre de nouveaux logements échantillonnés chaque mois était limité à une valeur établie d'avance appelée allocation de base. Dans l'ancien plan de sondage, il y avait une allocation de base distincte pour chaque grande catégorie d'échantillon dans chaque province. Les catégories étaient les suivantes :

Secteur autoreprésentatif, sans la base d'appartements (AR-norm.)

Secteur autoreprésentatif, base d'appartements (AR-app.)

Secteur non autoreprésentatif, région urbaine (NAR-urb.)

Secteur non autoreprésentatif, région rurale (NAR-rur.)

Secteurs spéciaux (SPE).

Les secteurs n'étaient pas tous soumis au programme de stabilisation. Les secteurs spéciaux de même que les petits secteurs autoreprésentatifs sans base d'appartements ne l'étaient pas. La définition des secteurs était intégrée au logiciel qui permettait de prélever l'échantillon de logements dans les grappes de l'EPA. Pour changer la définition d'un secteur de stabilisation, il fallait soumettre une proposition de modification de programme afin que le changement soit apporté à toutes les composantes du système.

Le retrait des logements se faisait selon l'algorithme suivant. Premièrement, les grappes hors-renouvellement et les grappes sous-pondérées en raison d'une forte croissance de la population étaient mises en évidence. Les grappes hors-renouvellement sont les grappes introduites un mois autre que celui indiqué par le numéro de renouvellement; elles proviennent principalement de la base ouverte d'appartements d'où elles ont été prélevées pour que les logements puissent être échantillonnés aussitôt que les grappes nouvellement ajoutées ont été choisies.

Les grappes sous-pondérées étaient prises en compte dans la taille de l'échantillon des secteurs de stabilisation, mais aucun des logements prélevés dans ces grappes ne pouvait être retiré et aucune de ces grappes ne se faisait attribuer un poids de stabilisation. Le poids de stabilisation à appliquer aux grappes restantes était calculé par la formule :

où:

W est le poids

N est le nombre de logements choisis au départ

B est l'allocation de base

C est le nombre de logements choisis dans les grappes sous-pondérées

and p is the province
 j is the stabilization area
 m is the rotation month

The number of dwellings to drop was thus (N-B). The drop was systematically applied to the file of (N-C) dwellings. If N was less than B then no stabilization was applied to that area for that month. The result was that some dwellings within some clusters were dropped. These dwellings were marked on the Form 02s with an 'S' to signal to the interviewer that they were part of the dropped sample.

The stabilization weights were not applied to sub-weighted clusters to avoid any increase in the sub-weight already present. They were not applied to off-rotation clusters since these clusters properly belong to a different birth rotation month.

Effects of Stabilization: As an illustration of the savings inherent in the stabilization program, from 1177 to 0984 over 75,000 dwellings were prevented from being interviewed, for a savings of \$2.7 million.

The stabilization program was also used by other survey vehicles. The Survey of Consumer Finances selected an independent set of households from the active set of clusters in the LFS. Stabilization allowed complete rotations to be reserved from different areas of the LFS frame while still limiting their sample size to a fixed total. Another major survey using the LFS frame was the Survey of Family Expenditures. The required allocations by different cities or city groups could not be incorporated in the existing stabilization program without major program modifications.

Redesign Changes: The redesign of the Labour Force Survey allowed some basic changes in the manner of sample size stabilization. In order to provide a more flexible program of stabilization the format for defining the base categories was modified to be more survey independent. Instead of fixed categories from the LFS frame, the stabilization areas were redefined in terms of ranges of Primary Sampling Units (PSUs). Special surveys such as the Survey of Consumer Finances and the Survey of Food/Family Expenditures could now also use the standard stabilization procedures by defining their own PSU ranges.

Not only the ranges of PSUs but also the basic theoretical weights to be applied at the weighting stage are now external inputs into the LFS system rather than being hard-coded into the various individual pieces of software. Thus, the user of the system has direct control without complex procedures involved in software changes. The formulation of the stabilization weights has remained the same.

et p est la province
 j est le secteur de stabilisation
 m est le mois de renouvellement

Le nombre de logements à retirer de l'échantillon était donc (N-B). Ce nombre était appliqué systématiquement au fichier des logements (N-C). Si N était inférieur à B , le secteur n'était pas stabilisé ce mois-là. Le résultat de cette opération était que certains logements, dans certaines grappes, étaient retirés. Pour signaler à l'intervieweur que ces logements faisaient partie des logements retirés de l'échantillon, on inscrivait un "S" sur la formule 02.

Les poids de stabilisation n'étaient pas appliqués aux grappes sous-pondérées afin que celles-ci ne soient pas sous-pondérées davantage. Ils n'étaient pas appliqués non plus aux grappes hors-renouvellement puisqu'en réalité ces grappes auraient dû être introduites un autre mois.

Effets de la stabilisation : Le tableau qui suit montre les économies que le programme de stabilisation a permis de réaliser. Entre 1177 et 0984, 75,000 logements ont été soustraits à l'interview, ce qui a engendré des économies de \$2.7 millions.

Le programme de stabilisation était également appliqué à d'autres enquêtes. Pour l'enquête sur les finances des consommateurs, un ensemble de ménages distinct était prélevé des grappes actives de l'EPA. Le programme de stabilisation permettait de réserver des groupes de renouvellement entiers dans différentes parties de la base de sondage de l'EPA tout en limitant la taille de l'échantillon à un nombre fixe d'unités. La seule autre grande enquête qui utilise la base de sondage de l'EPA est l'enquête sur les dépenses des familles. La répartition de l'échantillon entre certaines villes ou certains groupes de villes requise par ces enquêtes ne pouvait pas être incorporée dans le programme de stabilisation de l'époque, car cela aurait nécessité de trop grands changements.

Changements découlant du remaniement : Le remaniement de l'enquête sur la population active (EPA) a permis d'apporter des changements fondamentaux au programme de stabilisation de la taille de l'échantillon. Pour rendre ce programme plus souple, on a modifié les critères de définition des catégories de base afin que celles-ci ne soient pas les mêmes pour toutes les enquêtes. Au lieu de répartir la base de sondage de l'EPA en catégories fixes, on a redéfini les secteurs de stabilisation en termes de tranches d'unités primaires d'échantillonnage (UPE). Pour les enquêtes spéciales comme l'enquête sur les finances des consommateurs et l'enquête sur les dépenses alimentaires/les dépenses des familles, il est devenu possible de suivre les procédures de stabilisation normales tout en définissant des tranches d'UPE qui leur soient propres.

Autant les tranches d'UPE que les poids théoriques de base à appliquer à l'étape de la pondération sont à présent des données qu'on introduit dans le système de l'EPA plutôt que des éléments incorporés aux logiciels. Les utilisateurs du système peuvent donc prendre des décisions à cet égard qui n'entraînent pas des procédures complexes pour modifier les logiciels. La formulation des poids de stabilisation n'a pas changé.

Development of Base Figures

Sample Allocation: With the redesigned sample, changes to the definitions of base areas for stabilization were required. The departure from a self-weighting design meant that within a province the sample size was summed for major frame types (SR, NSR, SPE) with the same basic weight, i.e., with the same overall sampling rate. In addition, each Census Metropolitan Area (CMA) was given a separate stabilization area. A minimum rule was applied to the resulting breakdown of the sample. Each potential stabilization area had to have a minimum of 300 households in the sample. Any areas with less than 300 households were designated as unstabilized. All unstabilized areas within a province were assigned to the same stabilization area.

The remaining areas were numbered consecutively to form between two and sixteen areas per province. The allocated sample for each area became the initial base figure from which the final base figure is calculated. Note that CMAs had to have a minimum of 300 households in their sample for other reasons. Thus, each of them is now a separate stabilization area. In addition, no Special Area sample is stabilized and the Apartment frame is no longer separated from the regular SR frame.

Conversion to Dwellings: The final stage of sampling of the LFS is done six weeks prior to interviewing and is based on a list of dwellings for each selected cluster. Hence, the desired stabilization of household sample size must be converted into a base figure of dwellings for each stabilization area.

The major factors in this conversion are as follows:

Vacancy Rate: The major difference between households and dwellings is the set of responses: Vacant, Seasonal, Under Construction and Occupied By Persons not in the LFS. These comprise 10% to 15% of the total number of dwellings.

For each stabilization area (not including unstabilized areas of course) an estimate of the vacancy rate was calculated based on the existing off-line survey results for 1084 and 1184. Due to the size of the off-line redesign sample, the vacancy rate so calculated may be only a crude approximation of the actual rate obtained once the redesign implementation is complete.

Temporary Dockets: Sometimes not all clusters are listed in time for sampling. Also additional dwellings may be found once the interviewer starts the interview process. These late additions are not available to the stabilization program. Hence some compensation is made on the base figure to allow for these extra households after sampling. The additional dwellings average about 2.5% of the initial sample size.

Établissement des allocations de base

Répartition de l'échantillon : Une fois le plan de sondage remanié, il a fallu modifier la définition des secteurs de base aux fins de la stabilisation. Comme le plan n'était plus autopondéré, dans chaque province, on a additionné la taille de l'échantillon des principales catégories de bases de sondage (secteurs autoreprésentatifs, secteurs non autoreprésentatifs et secteurs spéciaux) qui avaient le même poids de base, c'est-à-dire la même fraction de sondage globale. On a convenu en outre que chaque région métropolitaine de recensement (RMR) constituait un secteur de stabilisation distinct. Il fallait que l'échantillon ainsi ventilé respecte un critère de taille minimum : l'échantillon de chaque secteur de stabilisation éventuel devrait compter au moins 300 ménages. Les secteurs où ce nombre n'était pas atteint étaient considérés comme des secteurs non stabilisés. Tous les secteurs non stabilisés d'une province ont été rattachés au même secteur de stabilisation.

On a ensuite donné des numéros consécutifs aux secteurs restants de manière à former de deux à seize secteurs par province. Le nombre d'unités dans l'échantillon attribué à chaque secteur est devenu l'allocation de base initiale à partir de laquelle l'allocation finale a été calculée. Il convient de noter que les RMR devaient avoir un échantillon d'au moins 300 ménages pour d'autres raisons aussi, de sorte que chacune d'elle est devenue d'office un secteur de stabilisation distinct. Par ailleurs, l'échantillon des secteurs spéciaux n'est pas stabilisé et la base d'appartements n'est plus séparée de la base normale des UAR.

Conversion en nombre de logements : On procède au dernier degré d'échantillonnage de l'EPA six semaines avant le début des interviews, qui sont faites à partir d'une liste de logements dans chaque grappe échantillonnée. Il faut donc convertir le nombre de ménages qu'on veut avoir dans l'échantillon après la stabilisation en une allocation de base de logements dans chaque secteur de stabilisation.

Les principaux facteurs qui interviennent dans cette conversion sont les suivants :

Taux de logements vacants : La principale différence entre le nombre de ménages et le nombre de logements est attribuable aux catégories "logement vacant", "logement saisonnier vacant", "logement en construction" et "logement occupé par des personnes ne devant pas être interviewées". Ces logements représentent environ 10 à 11% du nombre total de logements.

Dans chaque secteur de stabilisation (secteurs non stabilisés exclus, bien sûr), la valeur prévue du taux de logements vacants est calculée en fonction des résultats obtenus pour l'échantillon introduit en parallèle en 1084 et 1184. Compte tenu de la taille du nouvel échantillon introduit en parallèle, il se peut que le taux de logements vacants ainsi calculé ne soit qu'une vague approximation du taux obtenu une fois la mise en oeuvre du nouveau plan terminée.

Dossiers temporaires : Il arrive parfois que le listage de toutes les grappes ne soit pas fini avant l'échantillonnage, ou encore que l'intervieweur découvre de nouveaux logements après qu'il ait commencé les interviews. Les ajouts que cela représente échappent au programme de stabilisation. Il faut donc rajuster les allocations de base après l'échantillonnage pour tenir compte de ces ménages supplémentaires. Les logements ajoutés constituent en moyenne environ 2.5% de la taille initiale de l'échantillon.

Rotation Imbalance: In addition to the conversion of households to dwellings, some adjustments were made to the base figures for design differences in the allocation of sample by rotation. The anticipated number of dwellings selected in one rotation is not necessarily the same as that selected in other rotations. The sample selected each month is almost entirely drawn from one rotation. Hence, the base figure should reflect these differences.

The stabilization base figure is allocated to the six rotations of the LFS according to the proportion of the dwellings assigned to each rotation. Thus, a rotation with a larger expected sample take will have a larger base figure. Stabilization weights will thus be divided proportionally across rotations. It should be noted that the differences rarely set to more than 2% of the average base figure.

Relatively large differences are noted in the NSR design since the number of clusters assigned to a rotation was relatively small compared with the size of the whole stabilization area. Unfortunately, these differences do not remain constant since they are based on the current set of selected PSUs. As PSUs rotate and the balance of sample take by rotation changes the stabilization base must also be adjusted.

Clusters with high yields that are being sub-sampled are excluded from the actual stabilizing rotation but their sample counts are included in determining the number of dwellings to drop. Hence, some high stabilization weights can be expected in areas containing growth situations, especially where the total sample size of the area is relatively small (close to the 300 household minimum discussed earlier).

Unstabilized Areas: The total sample size for Canada was fixed at 51,675 households plus about 1,300 extra households due to a contract with the Alberta Government for a total of 52,975. Only 48,525 households are allocated to areas covered by the stabilization program. The rest are not stabilized at all (4,450 households or 8.5% of the total).

In effect this means that sample sizes in unstabilized areas are allowed to grow while stabilized areas are maintained relatively constant. Fortunately, unstabilized areas are mostly smaller cities with little prospect for large growth. Two years after implementation the yield in these areas totaled 4,510 households compared with the 4,450 allocated.

Répartition inégale entre les groupes de renouvellement : En plus de la conversion du nombre de ménages en nombre de logements, il faut aussi rajuster les allocations de base pour tenir compte des écarts attribuables au plan de sondage observés dans la répartition de l'échantillon entre les groupes de renouvellement. Le nombre prévu de logements choisis dans un groupe de renouvellement n'est pas nécessairement le même que celui qui est prélevé dans les autres groupes de renouvellement. L'échantillon prélevé chaque mois provient presque entièrement d'un groupe de renouvellement. L'allocation de base doit donc rendre compte des différences attribuables au plan de sondage.

L'allocation de base aux fins de la stabilisation est répartie entre les six groupes de renouvellement de l'EPA proportionnellement au nombre de logements attribué à chacun de ces groupes. Un groupe de renouvellement dont le rendement prévu de l'échantillon est élevé aura par conséquent une grande allocation de base. Les poids de stabilisation seront donc répartis proportionnellement entre les groupes de renouvellement. Il convient de noter que les différences représentent rarement plus de 2% de l'allocation de base moyenne.

On a relevé d'assez grands écarts dans les secteurs NAR, car le nombre de grappes attribué à un groupe de renouvellement y était relativement petit comparativement à la taille du secteur de stabilisation. Malheureusement, comme ces écarts sont liés aux UPE choisies, ils ne demeurent pas constants. Dès que les UPE sont remplacées et que le reste du rendement prévu de l'échantillon par groupe de renouvellement varie, il faut rajuster également l'allocation de base aux fins de la stabilisation.

Les grappes qui ont un rendement de l'échantillon élevé et qui sont sous-échantillonnées ne sont pas à proprement parler soumises au programme de stabilisation, mais on tient compte de la taille de leur échantillon au moment de déterminer le nombre de logements à laisser tomber. On peut donc s'attendre que certains poids de stabilisation soient élevés dans les secteurs où il y a des situations de croissance, particulièrement si la taille de l'échantillon global du secteur est relativement petite (près du minimum de 300 ménages mentionné plus haut).

Secteurs non stabilisés : L'échantillon pour l'ensemble du Canada avait été fixé à 51,675 ménages, plus environ 1,300 logements supplémentaires rajoutés en raison d'une entente conclue avec l'Alberta, ce qui en portait la taille totale à 52,975 ménages. Dans les faits, 48,525 ménages seulement sont attribués à des secteurs visés par le programme de stabilisation. Les autres en sont complètement exclus (4,450 ménages, ou 8.5% de l'échantillon total).

Cette situation signifie qu'on permet à la taille des échantillons dans les secteurs non stabilisés d'augmenter tandis qu'on maintient à un niveau relativement constant la taille des échantillons dans les secteurs stabilisés pour contrebalancer la croissance de l'échantillon global. Il se trouve qu'heureusement les secteurs non stabilisés sont pour la plupart de petites villes qui ont peu de chances de voir leur population augmenter. Deux ans après la mise en oeuvre du programme de stabilisation, le rendement de l'échantillon y était de 4,510 ménages, alors qu'on leur en avait attribué 4,550.

Implementation

Phase Out of Old Design: Due to the introduction scheme of the redesign the last three months of the old design involved a much reduced sample. Hence, the stabilization base figures for 1084 to 1284 inclusive had to be modified accordingly. Many areas that were completely on-line had no old sample in these months. Other areas had such small residual base figures that they were left unstabilized.

Phase In of New Design: During the first three months, from 1084 to 1284, the redesign sample was not stabilized at all. The implementation of the revised stabilization program started at 0185. As such, the full effect of stabilization was not felt until the June survey (0685), when all the rotations had gone through the new stabilization routines at sampling.

Results: The following tables are included to summarize the results of the above discussions. The final sample size in June of 1985 was 52,563 households compared with the allocation of 52,975, which indicates the imprecision inherent in trying to stabilize household size through the sampling routine six weeks prior to interviewing.

Mise en oeuvre du programme de stabilisation

Abandon de l'ancien plan : Compte tenu du mode d'introduction du nouveau plan de sondage, l'échantillon utilisé pendant les trois derniers mois où l'on se servait de l'ancien plan a été réduit de beaucoup. Il a donc fallu modifier en conséquence les allocations de base pour les mois 1084 à 1284 inclusivement. Dans de nombreuses régions où l'échantillon était entièrement introduit progressivement, aucune partie de l'échantillon n'a été prélevée selon l'ancien plan pendant cette période. Dans d'autres régions, les allocations de base résiduelles étaient si faibles que celles-ci n'ont pas été soumises à la stabilisation.

Introduction du nouvel échantillon : Pendant les trois premiers mois, soit de 1084 à 1284, l'échantillon prélevé selon le nouveau plan de sondage n'a pas été stabilisé du tout. La mise en oeuvre du nouveau programme de stabilisation a commencé au mois 0185. L'effet complet de la stabilisation ne s'est donc pas fait sentir avant l'enquête de juin (0685), après que tous les groupes de renouvellements aient été soumis au nouveau programme de stabilisation au moment de l'échantillonnage.

Résultats : Les tableaux qui suivent résument l'exposé que nous venons de faire. La taille finale de l'échantillon en juin 1985 était de 52,563 ménages, comparativement au nombre de 52,975 attribué, ce qui montre le manque de précision qui résulte de la stabilisation de l'échantillon des ménages au moment de l'échantillonnage, six semaines avant le début des interviews.

Table B-1: Actual Sample Sizes for October 1984 to June 1985

Month Mois	Old Design Total		New Design On-Line		New Design Off-Line		Grand Total	
	Total selon l'ancien plan		Echant. introd. progressiv./ Nouveau plan		Echant. introd. en parallèle/ Nouveau plan		Total	
	Dwg	Hhd	Dwg	Hhd	Dwg	Hhd	Dwg	Hhd
	Logements	Ménages	Logements	Ménages	Logements	Ménages	Logements	Ménages
1084	56484	50121	5920	5432	12799	11248	75203	66801
1184	50272	44390	11757	10833	17027	15048	79056	70271
1284	44140	38588	17501	16228	21178	18758	82819	73574
0185	12114	11030	48452	43770			60566	54800
0285	6107	5547	53934	48855			60041	54402
0385			59355	53883			59355	53883
0485			58964	53372			58964	53372
0585			58563	52945			58563	52945
0685			58244	52563			58244	52563

Table B-2: Stabilization Base Figures by Area Used for 0185 Survey

Area Secteur	Province									
	0	1	2	3	4	5	6	7	8	9
1	54	75	77	138	non	non	324	216	185	81
2	non	161	non	120	60	73	58	58	non	100
3	260		184	112	103	452	non	165	94	non
4	112		118	266	109	51	57	318	269	302
5			240	69	344	66	112	133	292	90
6				non	52	109	157	non	100	95
7					55	102			251	125
8					54	97			115	63
9					97	56			64	
10					106	54				
11					303	58				
12						62				
13						62				
14						293				
15						94				
16						108				

Tableau B-2: Allocations de base aux fins de la stabilisation pour l'enquête du mois 0185,

Table B-3: Identification of Stabilization Areas by Province

Tableau B-3: Identification des secteurs de stabilisation, par province

Province	Area	Province	Area
	Secteur		Secteur
Newfoundland – Terre-Neuve			
Carbonear, T.	1	La Baie	3
Gander, T.	1	Saint-Hyacinthe	3
Happy Valley, T-Goose Bay	1	Salaberry-De-Valleyfield	3
Labrador, C-Wabush	1	Shawinigan	3
Stephenville, T.	1	Sorel	3
Wabana-Bell Island	1	St-Jean	3
Windsor, T-Grand Falls	1	St-Jérôme	3
NSR-ER020 – NAR-RE020	2	Thetford Mines	3
SPECIAL AREAS – SECTEURS SPECIAUX	2	Victoriaville	3
NSR-ER010 – NAR-RE010	4	Ville de Québec	4
NSR-ER030 – NAR-RE030	4	NSR-ER411 – NAR-RE411	10
NSR-ER040 – NAR-RE040	4	NSR-ER412 – NAR-RE412	1
Channel-Port Aux Basques	2	NSR-ER420 – NAR-RE420	11
Corner Brook, C.	2	NSR-ER430 – NAR-RE430	11
Fortune, T-Grand Bank	2	NSR-ER440 – NAR-RE440	11
Marystown	2	NSR-ER450 – NAR-RE450	11
St. John's – Saint-Jean	3	NSR-ER461 – NAR-RE461	11
		NSR-ER463 – NAR-RE463	11
		NSR-ER464 – NAR-RE464	10
		NSR-ER470 – NAR-RE470	11
		NSR-ER480 – NAR-RE480	10
		SPECIAL AREAS – SECTEURS SPECIAUX	1
Prince Edward Island – Île-du-Prince-Édouard		Montreal PSU 46100-46219 – UPE 46100-46219	5
Charlottetown	1	Hull	6
Summerside	1	Chicoutimi-Jonquière	7
NSR-ER100 – NAR-RE100	2	Trois Rivières	8
		Montreal PSU 46220-46254 – UPE 46220-46254	9
Nova Scotia – Nouvelle-Écosse		Ontario	
Sydney Mines	1	Belleville-Trenton	1
Sydney-Glace Bay	1	Brantford	1
Bridgewater	2	Cornwall	1
Kentville	2	Kenora-Keewatin	1
Kingston-Greenwood	2	Kingston	1
Liverpool	2	Sarnia	1
New Glasgow	2	Guelph	2
Truro	2	Peterborough	2
Yarmouth	2	Sault Ste. Marie	2
SPECIAL AREAS – SECTEURS SPECIAUX	2	Toronto	3
NSR-ER210 – NAR-RE210	4	Oshawa	4
NSR-ER220 – NAR-RE220	4	Barrie	5
NSR-ER250 – NAR-RE250	4	Chatham	5
NSR-ER230 – NAR-RE230	5	Fort Erie	5
NSR-ER240 – NAR-RE240	5	Halton Hills	5
Halifax	3	Milton	5
		Newcastle	5
New Brunswick – Nouveau-Brunswick		St. Thomas	5
Barthust1	1	Timmins	5
Chatham/Newcastle	1	Woodstock	5
Edmundston	1	NSR-ER500 – NAR-RE500	14
Fredericton	1	NSR-ER510 – NAR-RE510	14
Grand Falls	1	NSR-ER520 – NAR-RE520	15
Dalhousie-Campbellton	2	NSR-ER530 – NAR-RE530	14
Moncton	2	NSR-ER540 – NAR-RE540	14
Oromocto	2	NSR-ER550 – NAR-RE550	14
Saint John	3	NSR-ER560 – NAR-RE560	14
NSR-ER310 – NAR-RE310	4	NSR-ER570 – NAR-RE570	14
NSR-ER320 – NAR-RE320	4	NSR-ER580 – NAR-RE580	16
NSR-ER330 – NAR-RE330	4	NSR-ER591 – NAR-RE591	14
NSR-ER340 – NAR-RE340	4	NSR-ER592 – NAR-RE592	15
NSR-ER350 – NAR-RE350	5	SPECIAL AREAS – SECTEURS SPECIAUX	1
SPECIAL AREAS – SECTEURS SPECIAUX	6	Eliot Lake	6
		Kirkland Lake	6
Quebec		North Bay	6
Baie-Comeau	1	Owen Sound	6
Gaspé	1	Stratford	6
Rimouski	1	Sudbury	6
Rouyn/Noranda	1	Ottawa	7
Sept Îles	1	Hamilton-Burlington	8
Val D'Or	1	Thunder Bay	9
Sherbrooke	2	Windsor	10
Alma	3	Kitchener-Waterloo	11
Drummondville	3	St. Catharines-Niagara	12
Grandby	3	London	13
Joliette	3		

Table B-3: Identification of Stabilization Areas by Province – Concluded

Province	Area Secteur
Manitoba	
Winnipeg	1
NSR-ER610 – NAR-RE610	4
NSR-ER620 – NAR-RE620	5
NSR-ER630 – NAR-RE630	6
NSR-ER640 – NAR-RE640	5
NSR-ER650 – NAR-RE650	6
NSR-ER660 – NAR-RE660	6
SPECIAL AREAS – SECTEURS SPECIAUX	3
Brandon	2
Dauphin	2
Flin Flon/The Pas	2
Thompson	2
Portage La Prairie	3
Selkirk	3

Alberta

Lethbridge	1
Medicine Hat	1
Red Deer	1
NSR-ER810 – NAR-RE810	6
NSR-ER820 – NAR-RE820	7
NSR-ER830 – NAR-RE830	6
NSR-ER840 – NAR-RE840	7
NSR-ER850 – NAR-RE850	8
NSR-ER860 – NAR-RE860	6
NSR-ER870 – NAR-RE870	9
NSR-ER880 – NAR-RE880	8
SPECIAL AREAS – SECTEURS SPECIAUX	2
Drumheller	2
Lloydminster	2
Fort McMurray	3
Grande Prairie	3
Calgary	4
Edmonton	5

Tableau B-3: Identification des secteurs de stabilisation, par province – fin

Province	Area Secteur
Saskatchewan	
Battleford	1
Estevan	1
Lloydminster	1
Melfort	1
Melville	1
Moose Jaw	1
North Battleford	1
Prince Albert	1
Swift Current	1
Wayburn	1
Yorkton	1
NSR-ER710 – NAR-RE710	4
NSR-ER720 – NAR-RE720	4
NSR-ER730 – NAR-RE730	4
NSR-ER740 – NAR-RE740	5
NSR-ER750 – NAR-RE750	4
SPECIAL AREAS – SECTEURS SPECIAUX	6
Regina	2
Saskatoon	3
British Columbia – Colombie-Britannique	
Cranbrook	1
Dawson Creek/Ft. St. John	1
Kimberley	1
Kitimat/Terrace	1
Prince Rupert	1
Kelowna	2
Mission/Matsqui	2
Penticton	2
Port Alberni	2
Powell River	2
Trail/Rossland	2
Vernon	2
Chilliwack	3
Kamloops	3
Nanaimo	3
Prince George	3
Vancouver	4
Victoria	5
NSR-ER910 – NAR-RE910	3
NSR-ER920 – NAR-RE920	6
NSR-ER930 – NAR-RE930	7
NSR-ER940 – NAR-RE940	6
NSR-ER950 – NAR-RE950	7
NSR-ER960 – NAR-RE960	7
NSR-ER970 – NAR-RE970	8
NSR-ER980 – NAR-RE980	8
SPECIAL AREAS – SECTEURS SPECIAUX	3

Future Implications

Regular Monitoring: Due to the sample related nature of estimating the stabilization base figures, a regular program of monitoring has been established to examine the validity of the assumed vacancy rate. Adjustments may be made every six months in order to keep the sample size consistently close to the desired number of households.

Since the stabilization areas are defined by province, any major changes in the sample sizes would affect the allocation of resources to those Regional Offices that split a province. A regular program of monitoring will spread these changes over many years as well as provide the capacity to report on current sample sizes by the Regional Offices.

Conséquences pour l'avenir

Contrôle régulier : Comme l'estimation des allocations de base aux fins de la stabilisation est étroitement liée à la taille de l'échantillon, un programme régulier de contrôle a été mis sur pied dont le but est d'examiner la validité du taux présumé de logements vacants. Des rajustements peuvent être faits tous les six mois dans le but de faire en sorte que la taille de l'échantillon se maintienne aux alentours de 52,975 ménages.

Les secteurs de stabilisation étant définis par province, tout changement important dans la taille des échantillons aurait une incidence sur l'attribution des ressources aux bureaux régionaux qui se partagent une province. Un programme régulier de contrôle aura pour effet d'étaler ces changements sur plusieurs années et permettra aussi de faire connaître la taille actuelle des échantillons par bureau régional.

Reallocation of Sample: It is reasonable to suppose that under certain conditions a reallocation of sample may occur during the life of the current design. The current sampling system can easily handle a change in the definition of the sample universe or in a simple reallocation of the available sample.

Possible scenarios include a reduction in the sample due to budgetary restraints. In this case the theoretical weight is increased for certain areas while the base figures are decreased correspondingly. In June 1986 an overall 10% sample size reduction was implemented. As the following table shows, the reduction was concentrated in major cities where the reliability of the sample was high.

Nouvelle répartition de l'échantillon : Il est raisonnable de supposer que, dans certaines conditions, on aura à procéder à une nouvelle répartition de l'échantillon pendant la période où le plan de sondage actuel sera en application. La méthode d'échantillonnage peut facilement se prêter à un changement de la définition de l'univers de l'échantillon ou à une simple nouvelle répartition de l'échantillon.

La réduction de l'échantillon en raison de contraintes budgétaires est une des éventualités. Si elle se présente, on augmente le poids théorique pour certains secteurs et on diminue les allocations de base en conséquence. En juin 1986, la taille globale de l'échantillon a été réduite de 10%. Comme le tableau qui suit le montre, cette réduction a été effectuée principalement dans les grandes villes où le degré de fiabilité de l'échantillon était élevé.

Table B-4: Sample Reduction (June 1986)

Tableau B-4: Réduction de l'échantillon (juin 1986)

City	Allocation Household	Percent Reduction	Sample Reduction
Ville	Nombre de ménages attribués	Réduction en pourcentage	Nombre de ménages soustraits
St. John's	637	25.00%	161
Sydney	322	20.00%	64
Halifax	1043	25.00%	261
Moncton	536	25.00%	134
Saint John - Saint-Jean	620	25.00%	155
Fredericton	348	20.00%	68
Quebec - Québec	609	10.00%	62
Montreal - Montréal	2499	33.33%	833
Hull	301	25.00%	75
Ottawa	589	25.00%	148
Toronto	2599	33.33%	866
Hamilton	560	10.00%	56
Winnipeg	1853	25.00%	463
Regina	962	25.00%	238
Moose Jaw	287	20.00%	56
Saskatoon	938	25.00%	232
Calgary	1484	23.00%	339
Edmonton	1631	23.00%	372
Vancouver	1704	33.33%	569
Victoria	554	10.00%	56

New Areas

At the moment the unstabilized areas do not present a problem for determining stabilization base figures for the other areas. If the sample becomes excessive in these areas there is the option of taking some of the PSU ranges currently unstabilized and creating a new area category. In effect this change implies a reduction in the minimum household size used as a criterion for assigning a separate stabilization area.

If the minimum household size for a stabilized area were lowered from 300 to 200, then about 1,790 extra households would enter the domain of stabilization. The inclusion of these areas in the stabilization would reduce the unstabilized allocations from 8.5% to 5.0% of the total sample size.

Nouveaux secteurs

Actuellement, les secteurs non stabilisés ne posent pas de problème en ce qui a trait à la détermination des allocations de base pour les autres secteurs aux fins de la stabilisation. Si l'échantillon devenait trop grand dans ces secteurs, il serait toujours possible de prendre certaines tranches d'UPE dans les secteurs non stabilisés et de créer une nouvelle catégorie de secteur. Cela reviendrait à faire baisser le nombre minimum de ménages qui sert de critère à la création d'un secteur de stabilisation.

Si on faisait passer le nombre minimum de ménages de 300 à 200, il y aurait environ 1,790 ménages de plus dans le programme de stabilisation. Le proportion de ménages non visés par le programme par rapport à l'échantillon global passerait alors de 8.5 à 5.0%.

The danger of including these new areas is that a large increase in sample size due to a growth cluster, for example, will produce a large stabilization weight to compensate. Instead of the stabilization program having a minor stabilizing effect the high weight would increase the variability of the data.

Il y aurait un danger à inclure ces nouveaux secteurs dans le programme de stabilisation, et c'est qu'une forte augmentation de la taille de l'échantillon, attribuable, par exemple, à la croissance de la population dans une grappe, aurait pour résultat un poids de stabilisation élevé qu'il faudrait contrebalancer. Au lieu que le programme ait un effet de stabilisation faible, le poids élevé augmenterait la variabilité des données.

Appendix C

Sample Phase-In

At the time of the preceding redesign, when the sample based on the 1971 Census was introduced, extensive tests were conducted through 1973 and 1974 to accommodate the major revisions to the survey. These tests culminated in a parallel run through all of 1975. The parallel run was a full parallel survey which used the then new sample and incorporated the revisions and innovations that had been developed. The phasing in of that sample was done gradually from 0574 through 1274, although the full sample expansion extended into 1976.

For the Labour Force Survey sample redesign introduction following the 1981 Census the prevailing consideration was timeliness, limited by quality assurance and cost implications. The requirement of a gradual introduction of the new sample extending over three or more months was respected by phasing in the new sample in steps from 1084 through 0385.

The requirement for a gradual phase-in of a new sample is dictated by three main factors:

(i) Rotation Group bias

The LFS sample consists of six equal panels identified by rotation numbers 1 to 6. Normally on any survey occasion, one panel is in the survey for the first occasion, another for the second, and so forth. For major labour force characteristics such as "unemployed" and "employed", there tends to be a small systematic difference among panels – a phenomenon known as rotation group bias. Unemployment rates are higher in the first month, for example, than in later months. It takes six months to introduce an LFS sample so that all of its panels are properly aged.

If the sample were introduced in less than six months, the expected value of survey estimates would differ initially from those of a properly aged sample. However, under the levels of rotation group bias observed in the old survey, the time period required to achieve an adequate approximation of a properly aged sample can be shortened to three or four months. During this time the new sample should not be used for deriving official estimates. The difference in estimates between panels of successive months decreases exponentially as the panel ages. By the third month a near stable picture appears.

Annexe C

Introduction de l'échantillon

Après le remaniement de l'avant dernier plan de sondage et l'introduction de l'échantillon basé sur le recensement de 1971, en 1973 et 1974, on a réalisé des essais très poussés pour vérifier l'effet des grands changements apportés à l'enquête, et ces essais ont abouti à la tenue d'un essai parallèle pendant toute l'année 1975. Ce qu'on a appelé l'essai parallèle était dans les faits une enquête à part entière auprès de ce qui était à l'époque le nouvel échantillon et dans laquelle on avait incorporé toutes les révisions et innovations effectuées. L'introduction de cet échantillon s'est faite progressivement entre le mois 0574 et le mois 1274, mais ce n'est qu'en 1976 que l'échantillon élargi a été entièrement en place.

Lorsqu'on a commencé le remaniement de l'enquête sur la population active après le recensement de 1981, le principal souci était l'actualité des données, qu'on ne devait chercher à obtenir ni aux dépens de la qualité des données ni à n'importe quel coût. Il fallait aussi que le nouvel échantillon soit introduit graduellement sur une période de trois mois ou plus, ce qui a été fait en procédant par étape du mois 1084 au mois 0385.

C'est pour tenir compte de trois facteurs qu'on doit introduire un nouvel échantillon graduellement :

(i) Biais de renouvellement

L'échantillon de l'EPA est constitué de six panels de taille égale portant les numéros de renouvellement 1 à 6. Normalement, chaque fois que l'enquête a lieu, il y a un des panels qui participe à l'enquête pour la première fois, un autre, pour la deuxième, et ainsi de suite. Pour les principales caractéristiques de la population active, comme le fait d'être en chômage ou d'avoir un emploi, la valeur prévue varie selon le panel. C'est ce qu'on appelle le biais de renouvellement. Les taux de chomages sont plus élevés le premier mois, par exemple, que pendant les derniers mois. Il faut compter six mois pour introduire un échantillon de l'EPA de manière à ce que tous les panels aient l'ancienneté idéale.

Si on introduisait l'échantillon en moins de six mois, la valeur prévue des estimations de l'enquête serait différente au départ de celle qu'on obtiendrait si l'échantillon avait la bonne composition du point de vue de l'ancienneté. Cependant, compte tenu du biais de renouvellement observé selon l'ancien plan, il serait possible de réduire à trois ou quatre mois la période nécessaire pour se rapprocher des résultats obtenus avec un échantillon ayant le profil d'ancienneté idéale, mais il ne faudrait pas se baser sur les données recueillies auprès du nouvel échantillon pour calculer les estimations officielles. L'écart entre les estimations obtenues pour les différents panels pendant des mois consécutifs diminue de façon exponentielle au fur et à mesure que l'échantillon se rapproche du profil d'ancienneté idéal. Dès le troisième mois, la situation se stabilise.

(ii) New Interviewer Effect

A consequence of the new sample being selected independently of the old was a requirement to hire new interviewers in Non Self-Representing Areas wherever the location of the sample changed. Past studies have shown that the work of new interviewers differs from that of experienced interviewers in initial months, although comparability of work is achieved after approximately three months. The differences relate both to operational measures, such as non response and edit failure rates, and to labour force characteristics. Hence, wherever a sizable proportion of new interviewers is required, the redesigned sample should not be used for official estimates until after a period of stabilization.

(ii) Effet de l'inexpérience des nouveaux intervieweurs

Lorsque le nouvel échantillon est prélevé de façon indépendante par rapport à l'ancien, il faut embaucher de nouveaux intervieweurs dans les secteurs non autoreprésentatifs partout où l'emplacement des unités échantillonnées n'est plus le même. Des études réalisées par le passé ont montré que, pendant les premiers mois de l'enquête, il y a des différences entre le travail accompli par les nouveaux intervieweurs et le travail accompli par les intervieweurs ayant de l'expérience, mais qu'au bout de trois mois les résultats sont comparables. Les différences observées se situent aussi bien sur le plan opérationnel (taux de non réponse et rejets à la vérification) que sur le plan des caractéristiques de la population active. Par conséquent, lorsqu'il est nécessaire d'embaucher une proportion importante de nouveaux intervieweurs, il ne faut pas diffuser d'estimations officielles calculées d'après les données recueillies auprès de l'échantillon prélevé selon le nouveau plan de sondage tant qu'une période de stabilisation ne s'est pas écoulée.

(iii) Field Capacity

When they are first introduced to the sample, dwellings require more time to interview; hence data quality related aspects (i) and (ii) aside, interviewers could not cope with introducing an entirely new sample in a single month. Rather, this work has to be spread over several months.

(iii) Possibilités du personnel sur le terrain

Lorsque les ménages participent à l'enquête pour la première fois, il faut prévoir plus de temps pour l'interview. Par conséquent, même si l'on faisait abstraction des facteurs (i) et (ii), les intervieweurs ne seraient pas en mesure de faire en un mois les interviews auprès d'un échantillon entièrement composé de nouveaux ménages. Il faut que ce travail soit étalé sur plusieurs mois.

Three alternative phase-in strategies considered were as follows:

Trois stratégies pour l'introduction du nouvel échantillon ont été examinées :

Strategy 1: On-line implementation

Under this strategy, the new sample would be introduced month-by-month by simply replacing the dwellings from the existing design with new ones from the redesign in the appropriate rotation. Operational considerations, namely the pace at which the work was proceeding, dictated that the earliest month in which such a phase-in could begin was October 1984. The pattern of interviewing by rotation is given below.

Stratégie 1 : Introduction progressive

Selon cette stratégie, le nouvel échantillon serait introduit un mois à la fois en remplaçant simplement les logements prélevés selon l'ancien plan par des logements prélevés selon le nouveau plan en respectant les numéros de renouvellement. Compte tenu de critères d'ordre opérationnel, plus précisément la vitesse à laquelle le travail avançait, il a été convenu que l'introduction progressive ne pouvait pas commencer avant le mois d'octobre 1984. Le mode de fonctionnement serait le suivant :

Rotation Number	Numéro de renouvellement	
Date	New Sample Nouvel échantillon	Old Sample Ancien échantillon
October 1984 - Octobre 1984	4	5, 6, 1, 2, 3
November 1984 - Novembre 1984	4, 5	6, 1, 2, 3
December 1984 - Décembre 1984	4, 5, 6	1, 2, 3
January 1985 - Janvier 1985	4, 5, 6, 1	2, 3
February - Février 1985	4, 5, 6, 1, 2	3
March - Mars 1985	4, 5, 6, 1, 2, 3	-

During this period, official estimates would be based on weighted averages of the separate estimates for the old and new sample, e.g.,

Oct. 84 official estimate
= 5/6 estimate from old sample
+ 1/6 estimate from new sample.

Strategy 2: Off-line implementation

In this strategy, the new sample would be allowed to build up in parallel with the existing sample. During this time period official estimates would be based entirely on the old sample. Once the new sample had been adequately aged to prevent rotation group bias and new interviewer effects, the old sample would be dropped and the new sample would continue. This strategy is represented below:

Rotation Number

Date

September 1984 – Septembre 1984
October 1984 – Octobre 1984
November 1984 – Novembre 1984
December 1984 – Décembre 1984
January 1985 – Janvier 1985

Pendant cette période, les estimations officielles seraient basées sur des moyennes pondérées des estimations se rapportant à l'ancien et au nouvel échantillons. Par exemple :

Estimations officielles pour oct. 84
= 5/6 des estimations d'après l'ancien plan
+ 1/6 des estimations selon le nouveau plan.

Stratégie 2 : Introduction en parallèle

Selon cette stratégie, le nombre de ménages du nouvel échantillon interrogés augmenterait alors que, parallèlement, l'enquête se poursuivrait auprès de l'ancien échantillon. Pendant ce temps, les estimations officielles seraient entièrement fondées sur les données recueillies auprès de l'ancien échantillon. Une fois que le nouvel échantillon aurait un profil d'ancienneté suffisamment bon pour qu'on n'ait à craindre ni le biais de renouvellement ni l'effet de l'inexpérience des nouveaux intervieweurs, on abandonnerait l'ancien échantillon et on poursuivrait avec le nouveau. Cette stratégie serait représentée de la manière suivante :

Numero de renouvellement

New Sample

Old Sample

Nouvel échantillon

Ancien échantillon

-	4 5 6 1 2 3
2* 3* 4*	5 6 1 2 3 4
2* 3* 4* 5*	6 1 2 3 4 5
2* 3* 4* 5* 6*	1 2 3 4 5 6
2 3 4 5 6 1	

* Sample not used for official estimates – Échantillon sur lequel on ne baserait pas les estimations officielles

Strategy 3: Combined Strategy

Under this strategy all areas would be designated as either on-line or off-line. The on-line method as described would be used in those provinces where Economic Region boundaries had not changed between 1971 and 1981 and where most of the old interviewers could be retained for the new sample.

In those provinces where Economic Region boundaries had changed or where a significant number of new interviewers was anticipated, the introduction would be off-line according to the schedule as shown for strategy 2. To further minimize the amount of off-line sample in these provinces, core portions of SR areas unaffected by boundary changes would also be introduced on-line.

Figure C-1 describes the composition of the Labour Force Survey sample by rotation group each month during the phase-in period for the two different types of areas.

Stratégie 3 : Stratégie mixte

Selon cette stratégie, les secteurs seraient répartis en deux catégories : ceux où l'introduction de l'échantillon se ferait progressivement et ceux où elle se ferait en parallèle. La méthode d'introduction progressive de l'échantillon décrite plus haut serait utilisée dans les provinces où les limites des régions économiques n'ont pas changé entre 1971 et 1981 et où l'on pourrait garder la plupart des anciens intervieweurs pour le nouvel échantillon.

Dans les provinces où les limites des régions économiques ont changé ou où l'on prévoit qu'il faudra embaucher un grand nombre de nouveaux intervieweurs, l'introduction se ferait en parallèle selon le calendrier présenté pour illustrer la stratégie 2. Pour réduire encore davantage la taille de l'échantillon introduit en parallèle dans ces provinces, l'échantillon serait introduit progressivement dans les parties centrales des secteurs AR qui n'ont pas été touchées par les changements de limites.

La figure C-1 décrit la composition de l'échantillon de l'enquête sur la population active par groupe de renouvellement à chaque mois de la période d'introduction selon les deux catégories de secteurs.

Figure C-1

**Composition of Sample by Rotation Group
Under the Combined Strategy**

	Rotation	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March
	Gr. de renouv.				Déc.	Jan.	Fév.	Mars
On-line areas – Secteurs d'introduction progressive	1	O	O	O	O	N	N	N
	2	O	O	O	O	O	N	N
	3	O	O	O	O	O	O	N
	4	O	N	N	N	N	N	N
	5	O	O	N	N	N	N	N
	6	O	O	O	N	N	N	N
Off-lines areas – Secteurs d'introduction en parallèle	1	O	O	O	O	N	N	N
	2	O	Oa	Oa	Oa	N	N	N
	3	O	Oa	Oa	Oa	N	N	N
	4	O	Oa	Oa	Oa	N	N	N
	5	O	O	Oa	Oa	N	N	N
	6	O	O	O	Oa	N	N	N

O = old sample – A = ancien échantillon

N = new sample for official estimates purposes – N = nouvel échantillon, servant aux estimations officielles.

a = advanced new sample (not used for official estimates) – n = nouvel échantillon avancé (ne servant pas aux estimations officielles)

Rationale for choice of Phase-in Strategy

The phase-in strategy adopted was number 3, the combined strategy. This section traces the rationale leading to its choice by examining the strengths and weaknesses of the on- line and off-line strategies.

On-line Strategy:

Advantages

- (i) It retains the normal panel structure of the LFS sample, so that there are no problems with rotation group bias.
- (ii) Its cost is minimal. Since the old sample is dropped rotation by rotation as the new sample is introduced, there is no doubling of new and old samples.

Disadvantages

- (i) Subprovincial Estimation in the Face of Boundary Changes

Problems arise for subprovincial areas, such as Economic Regions or CMAs whose boundaries have changed, since estimates for either old or new boundaries could not be produced from either sample by using standard estimation techniques.

Figure C-1

**Composition de l'échantillon par groupe de
renouvellement selon la stratégie mixte**

Raisons expliquant le choix de stratégie

La stratégie adoptée pour l'introduction de l'échantillon est la stratégie 3, soit la stratégie mixte. Dans cette section, nous allons examiner les raisons qui ont amené les responsables de l'enquête à la choisir en nous penchant tour à tour sur les avantages et les inconvénients de la stratégie d'introduction progressive et de la stratégie d'introduction en parallèle.

Stratégie d'introduction progressive

Avantages

- (i) Elle respecte la structure normale des panels de l'échantillon, de sorte qu'il n'y a pas de problème de biais de renouvellement.
- (ii) Son coût est minime. Comme on laisse tomber l'ancien échantillon groupe de renouvellement par groupe de renouvellement, au fur et à mesure que le nouvel échantillon est introduit, on n'utilise jamais l'ancien et le nouvel échantillons en même temps.

Inconvénients

- (i) Estimations infraprovinciales là où les limites on changé

On se heurte à des difficultés dans les régions infraprovinciales comme les régions économiques et les RMR où les limites ont changé, car il n'est possible de produire des estimations ni selon les anciennes limites si selon les nouvelles en se basant sur l'un ou l'autre des échantillons au moyen des techniques d'estimation normales.

(ii) Lack of Stabilization Period for New Interviewers

In the NSR portion of some provinces, it was anticipated that a high proportion of new interviewers would be required. Under the on-line strategy, the use of new sample for estimation purposes from the outset would create the possibility of a significant new interviewer effect on the relevant estimates.

(ii) Pas de période de stabilisation pour les nouveaux intervieweurs

Il était prévu qu'on aurait besoin d'une grande proportion de nouveaux intervieweurs dans la partie NAR de certaines provinces. Lorsqu'on applique la stratégie d'introduction progressive, si on se fonde dès le départ sur les données recueillies auprès du nouvel échantillon pour produire les estimations, l'effet de l'inexpérience des nouveaux intervieweurs risque d'entacher fortement ces dernières.

(iii) Small Initial Assignments for Interviewers

Where new interviewers are required, initial assignment sizes might cause turnover.

(iii) Petites tâches initiales pour les intervieweurs

Lorsqu'il faut embaucher de nouveaux intervieweurs, si la tâche initiale est trop grande, le roulement de personnel risque d'être élevé.

There were no satisfactory solutions to the first two disadvantages, the most important ones.

On n'a pas trouvé de solution satisfaisante aux deux premiers, et plus graves, inconvénients.

Off-line Strategy

The off-line strategy overcomes the three problems associated with the on-line strategy. Moreover, rotation group bias from estimates based on a correctly aged sample would be negligible. In contrast, there are drawbacks in terms of cost and burden on field staff when old interviewers are retained for the new sample, since there is no lessening of work related to the old sample.

Stratégie d'introduction en parallèle

La stratégie d'introduction de l'échantillon en parallèle règle les deux problèmes associés à la stratégie d'introduction progressive. En outre, le biais de renouvellement est négligeable si les estimations sont fondées sur un échantillon ayant un profil d'ancienneté suffisamment bon. En revanche, elle a deux faiblesses : les coûts sont élevés, de même que le fardeau imposé au personnel sur le terrain dans les régions où l'on demande aux intervieweurs déjà en place de mener l'enquête auprès du nouvel échantillon tout en continuant leur travail auprès de l'ancien.

Advantages**Avantages**

(i) New Interviewer Effect

The new sample builds up for three months before it is used. This is deemed a sufficient break-in period for new interviewers.

(i) Effet de l'inexpérience des nouveaux intervieweurs

On laisse le nouvel échantillon s'accroître pendant trois mois avant de se servir des données recueillies, ce qui donne aux nouveaux intervieweurs une période de rodage qu'on estime suffisante.

(ii) New Interviewer Workloads

New interviewers are started with three-sixths of a full assignment comprising only first, hence longer, interviews. Second and third month interviewers receive four-sixths and five-sixths of a normal assignment.

(ii) Charge de travail des nouveaux intervieweurs

On attribue aux nouveaux intervieweurs les 3/6 d'une tâche complète pour commencer, et il s'agit de premières interviews qui donc prennent plus de temps. Le deuxième mois, on leur attribue les 4/5 d'une tâche normale et le troisième, les 5/6.

(iii) Impact of Rotation Group Bias

Calculations by Ghangurde (1982) show that rotation group biases from a correctly aged sample are small enough not to cause problems.

(iii) Conséquences du biais de renouvellement

Selon les calculs effectués par Ghangurde (1982), le biais de renouvellement est assez faible lorsque l'échantillon a un profil d'ancienneté suffisamment bon pour que cela ne pose pas de problème.

Disadvantages**Inconvénients**

(i) Old interviewers retained for the new sample are required for both samples during the phase-in period, feasible only if survey weeks in off-line areas are different.

(i) Les intervieweurs déjà en place auxquels on attribue des ménages dans le nouvel échantillon doivent travailler sur les deux échantillons pendant la période d'introduction, ce qui n'est possible que si la semaine d'enquête n'est pas la même dans les secteurs où l'échantillon est introduit en parallèle.

- (ii) The cost of extra interviews for off-line areas is estimated at \$1.1 million compared with the negligible cost of the on-line scheme.

- (ii) Le coût estimé des interviews additionnelles lorsque l'échantillon est introduit en parallèle est de 1.1 million de dollars, alors qu'il est négligeable lorsque l'échantillon est introduit progressivement.

Combined Strategy

The combined strategy was adopted because it resorts to the on-line strategy wherever it is not problematic. This is done on the grounds that it is preferable in terms of both cost and field operations. Wherever the on-line method causes problems in areas affected by boundary changes and where a high proportion of new interviewers is required, the off-line method provides an effective solution.

Under the combined strategy, the cost of extra interviews is estimated at \$350,000 and the effect on rotation group biases is minimal.

Strategy Application

The criteria that determined how each sample component was phased-in were boundary changes and the anticipated need for new interviewers. A revision of Economic Region boundaries was requested by Alberta, Ontario and Quebec. The new ERs in Ontario and Alberta consisted of different groupings of Census Divisions, whereas in Quebec the new ERs were defined using provincially defined small areas known as Municipalités Régionales de Comté (MRC). Due to the lack of comparability of ERs between the old design and the redesign, these provinces were designated to be introduced off-line.

Because of geographical difficulties in British Columbia and Newfoundland the new sample required a significant proportion of new interviewers. Hence, due to the above factors, Newfoundland, Quebec, Ontario, Alberta and British Columbia were designated as off-line and Prince Edward Island, Nova Scotia, New Brunswick, Manitoba and Saskatchewan were designated as on-line.

SRUs in Off-line Provinces

All SRUs for which the boundaries remained unchanged by the redesign were introduced on-line. In those SRUs where there was annexation or deletion, core areas that were collections of complete 1971 strata unaffected by boundary changes were introduced on-line. The 1981 strata were designed to respect the boundary of the so defined core.

Exceptions in On-line Provinces

Annexations to Census Metropolitan Areas in these provinces made it necessary to modify the phase-in stage. In Nova Scotia a significant portion of NSR from ER 25 was annexed to the CMA of Halifax. A core

Stratégie mixte

La stratégie mixte a été adoptée parce qu'elle permet d'avoir recours à l'introduction progressive, qui est préférable du point de vue des coûts et du travail sur le terrain, partout où cela ne pose pas de problème. Là où la méthode d'introduction progressive pose des problèmes, soit parce que les limites des unités ont changé ou parce qu'elle nécessite qu'on embauche un trop grand nombre de nouveaux intervieweurs, la méthode d'introduction en parallèle constitue une bonne solution de rechange.

Avec la stratégie mixte, le coût estimé des interviews additionnelles est de \$350,000 et les biais de renouvellement sont minimes.

Application de la stratégie

Les critères sur lesquels on s'est fondé pour déterminer comment introduire chaque partie de l'échantillon étaient les changements apportés aux limites et la nécessité prévue d'embaucher de nouveaux intervieweurs. L'Alberta, l'Ontario et le Québec avaient demandé que les limites des régions économiques soient modifiées. Les nouvelles RE en Ontario et en Alberta ont été constituées de nouveaux groupements de divisions de recensement, tandis qu'au Québec on est parti de petites régions définies par la province et appelées des municipalités régionales de comté (MRC). Comme les RE du nouveau plan n'étaient plus comparables à celles de l'ancien, il a été convenu que l'échantillon dans ces provinces serait introduit en parallèle.

En Colombie-Britannique et à Terre-Neuve, il fallait, en raison de difficultés d'ordre géographique, embaucher une proportion considérable de nouveaux intervieweurs pour le nouvel échantillon. Il a donc été décidé que l'échantillon de ces deux provinces aussi serait introduit en parallèle, ce qui laissait l'Île-du-Prince-Édouard, la Nouvelle-Écosse, le Nouveau-Brunswick, le Manitoba et la Saskatchewan comme provinces où l'échantillon serait introduit progressivement.

UAR dans les provinces où l'échantillon a été introduit en parallèle

Dans toutes les UAR où les limites n'avaient pas changé avec le nouveau plan de sondage, l'échantillon a été introduit progressivement. Pour ce qui est de celles où des parties avaient été ajoutées ou supprimées, on a procédé à l'introduction progressive de l'échantillon dans les parties centrales constituées de collections de strates de 1971 complètes non touchées par les changements de limites. Les strates de 1981 ont été formées de manière à respecter les limites des parties centrales ainsi définies.

Exceptions dans les provinces où l'échantillon a été introduit progressivement

Il a fallu modifier le plan d'introduction de l'échantillon dans les provinces où il y a eu des annexions aux régions métropolitaines de recensement. En Nouvelle-Écosse, une grande partie des secteurs NAR de la RE 25 a été annexée à

area consisting of the original CMA was introduced on-line. The remainder of the ER, including the annexed parts to become SR in the new design, was introduced off-line.

Similarly a part of NSR from ER 61 in Manitoba had been added to the CMA of Winnipeg. Again, the core part of Winnipeg corresponding to the old CMA was on-line and all of ER 61 was off-line. Changes were also made to the CMA of Saskatoon in ER 73. All of this ER was introduced off-line except the apartment sample in Saskatoon.

Apartment Sample

In the redesign, 18 SR areas have separate apartment sample strata. Except for Hull, the SRUs were introduced on-line and therefore so was the apartment sample. In Saskatoon, which was to be off-line, the apartment sample was introduced on-line, since the frame is concentrated in the core. In Hull, the apartment sample was set up for the first time as part of the redesign and was introduced off-line, along with the SRU.

Allocation of Sample

As a result of applying the mixed strategy, just over 58% of the sample was phased in on-line. Table C-2 presents the new sample as a percentage of the total sample used for official estimates by province for each month during the phase-in period.

Table C-2

Percentage of New Sample by Month and Province during Redesign Sample Phase-in

	October	November	December	January	February	March
	Octobre	Novembre	Décembre	Janvier	Février	Mars
Newfoundland – Terre-Neuve	3.3	6.8	10.3	91.9	96.0	100.0
Prince Edward Island – Île-du-Prince-Édouard	17.4	34.5	48.8	67.9	84.1	100.0
Nova Scotia – Nouvelle-Écosse	15.5	30.4	45.7	67.9	83.9	100.0
New Brunswick – Nouveau-Brunswick	16.5	33.2	50.1	67.0	83.6	100.0
Quebec	6.0	12.2	18.4	86.9	93.3	100.0
Ontario	9.0	17.9	26.6	83.4	91.8	100.0
Manitoba	13.3	26.8	40.9	66.4	82.3	100.0
Saskatchewan	11.6	23.7	36.0	73.0	86.5	100.0
Alberta	6.2	12.1	18.2	88.1	94.0	100.0
British Columbia – Colombie-Britannique	6.1	12.2	18.4	84.9	92.8	100.0

la RMR de Halifax. Dans la partie centrale constituée de la RMR d'origine, l'échantillon a été introduit progressivement. Dans le reste de la RE, y compris les parties annexées qui allaient devenir AR dans le nouveau plan, l'échantillon a été introduit en parallèle.

La même chose s'est passée au Manitoba : une partie des secteurs NAR de la RE 61 a été ajoutée à la RMR de Winnipeg. L'échantillon a été introduit progressivement dans la partie de Winnipeg correspondant à l'ancienne RMR et en parallèle dans l'ensemble de la RE 61. Des changements ont été apportés à la RMR de Saskatoon, dans la RE 73. L'échantillon a été introduit en parallèle dans toute cette RE sauf dans la partie qui avait été prélevée dans la base d'appartements de Saskatoon.

Echantillon prélevé dans la base d'appartements

Selon le nouveau plan de sondage, il y a des strates formées à partir d'une base d'appartements dans 18 UAR. Partout sauf à Hull, l'échantillon a été introduit progressivement, et donc la partie prélevée dans la base d'appartements aussi. A Saskatoon, où l'échantillon devait être introduit en parallèle, la partie prélevée dans la base d'appartement a été introduite progressivement, car elle était concentrée dans le centre de la ville. A Hull, la base d'appartements a été créée dans le cadre du remaniement du plan de sondage, et l'échantillon qui y a été prélevé a été introduit en parallèle, comme dans le reste de l'UAR.

Répartition de l'échantillon

L'adoption de la stratégie mixte a eu pour conséquence qu'un peu plus de 58% de l'échantillon a été introduit progressivement. Le tableau C-2 indique la proportion de l'échantillon utilisé pour les estimations officielles qui était représentée par le nouvel échantillon, selon le mois et la province, pendant l'introduction de l'échantillon.

Tableau C-2

Pourcentage représenté par le nouvel échantillon, selon le mois et la province, pendant l'introduction de l'échantillon prélevé d'après le nouveau plan de sondage

Sample Size Stabilization During Phase-In

Sampling rates for the LFS are established to yield a desired sample size at the outset of the design. Over time, however, continued application of the original rates would result in a 2%-3% annual increase in the sample size, since the universe of dwellings is growing at this rate.

Sample size stabilization is a step in the sample selection process to prevent such growth in the sample size. Under this step, the sample of newly selected dwellings each month is systematically cut back to the desired or "base" sample size, and corresponding factors to increase the weights of non-dropped dwellings are introduced (Drew, 1977). In the old design, the sample was stabilized separately for SR apartments, regular SR, NSR urban and NSR rural within each province.

From October to December 1984, sample stabilization was restricted to new dwellings selected under the old design in off-line areas. Hence, dwellings from the new design were split off prior to stabilization. The base sample sizes were reduced by the proportion of the old sample that was off-line in each province/area considered in stabilization.

A revised stabilization program was introduced for January 1985. No sample was selected from the old design at this time although two rotations were still phasing out. January was thus the first month where the full sample selection was derived only from the new sample (i.e., Rotation 1). Hence, since only the births each month are stabilized the new sample became completely stabilized only in June 1985.

The new stabilization features a more flexible definition of the area to be stabilized in terms of ranges of PSUs. Typically, CMAs and SR and NSR portions of Economic Regions form stabilization areas. Special surveys especially can now specify their own set of stabilization areas.

Base figures for the redesign sample were based on expected households under the new sample allocation. These figures were determined from 1981 Census data together with current survey estimates of the proportion of vacants in each stabilized area. The vacancy rate was required to convert households into dwellings since the sample is in terms of dwellings. The off-line new sample for October and November was used to finalize these base figures prior to January survey sampling in the first week of December.

Stabilisation de la taille de l'échantillon pendant l'introduction de celui-ci

Les fractions de sondage pour l'EPA sont établies au début de l'élaboration du plan de sondage et visent à donner un échantillon d'une taille voulue. Cependant, si l'on continuait d'appliquer les mêmes fractions, la taille de l'échantillon s'accroîtrait de 2 à 3% par an, car c'est là le rythme de croissance de l'univers des logements.

La stabilisation de la taille de l'échantillon est une étape de l'échantillonnage dont le but est de prévenir la croissance de l'échantillon. Elle consiste à réduire systématiquement le nombre de nouveaux logements échantillonnés chaque mois pour ramener ce nombre à une allocation de base et à introduire des facteurs ayant pour objet d'augmenter en conséquence le poids des logements demeurant dans l'échantillon (Drew 1977). Dans l'ancien plan de sondage, on procédait séparément à la stabilisation des échantillons prélevés dans les UAR avec base d'appartements, dans les UAR normales, dans les régions urbaines, des secteurs NAR et dans les régions rurales des secteurs NAR de chaque province.

D'octobre à décembre 1984, on a limité la stabilisation de l'échantillon aux nouveaux logements prélevés selon l'ancien plan de sondage dans les régions où l'échantillon était introduit en parallèle. Les logements prélevés selon le nouveau plan de sondage ont donc été retirés avant la stabilisation. Dans chaque province ou secteur soumis à la stabilisation, la taille de base des échantillons a été diminuée proportionnellement au pourcentage représenté par l'ancien échantillon dans les secteurs où l'échantillon a été introduit en parallèle.

On a introduit un nouveau programme de stabilisation en janvier 1985. Aucune partie de l'échantillon n'était prélevée selon l'ancien plan de sondage à ce moment-là, quoiqu'il restait deux groupes de renouvellement qui terminaient leur cycle. Le mois de janvier était donc le premier mois où tout l'échantillon prélevé était uniquement selon le nouveau plan (autrement dit, tout le groupe de renouvellement 1). Comme seulement les nouveaux ménages du mois sont soumis à la stabilisation, ce n'est qu'en janvier 1985 que le nouvel échantillon a été entièrement stabilisé.

La définition de ce qui constitue un secteur de stabilisation en termes de tranches de codes d'UPE est plus souple dans le nouveau programme de stabilisation. En général, les RMR ainsi que les parties AR et NAR des régions économiques sont des secteurs de stabilisation. Il est notamment possible à présent de décider quels secteurs seront des secteurs de stabilisation aux fins des enquêtes spéciales.

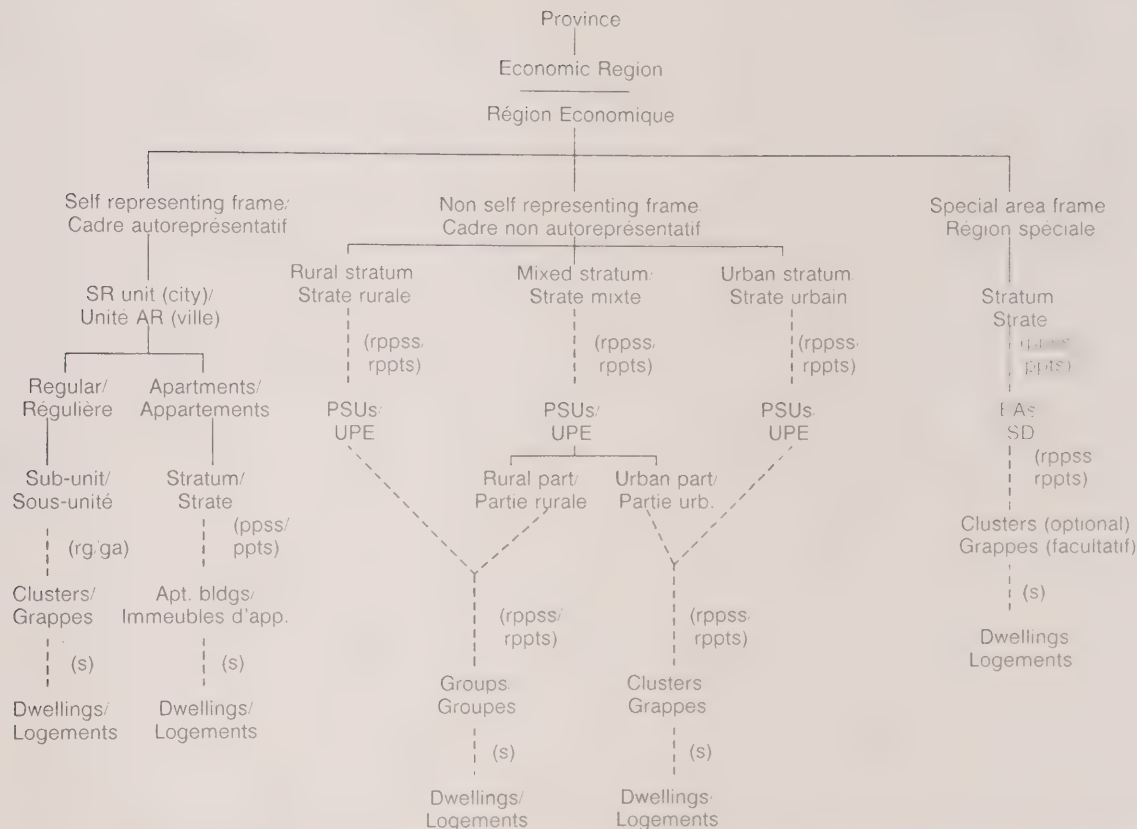
Les allocations de base pour l'échantillon prélevé selon le nouveau plan de sondage ont été établies d'après le nombre prévu de ménages selon la nouvelle répartition, lequel est calculé au moyen des chiffres du recensement de 1981 et de l'estimation de la proportion de logements vacants dans chaque secteur de stabilisation d'après les résultats de la dernière enquête. Il est nécessaire de connaître le taux de logements vacants pour convertir le nombre de ménages en nombre de logements, l'échantillon étant un échantillon de logements. On s'est servi des résultats obtenus pour l'échantillon introduit en parallèle en octobre et novembre pour apporter les dernières corrections nécessaires aux allocations de base avant le prélèvement de l'échantillon pour l'enquête de janvier, la première semaine de décembre.

Appendix D

Diagram of the LFS Frame

Annexe D

Diagramme de la base de l'EPA



Abbreviations:

pps – probability proportional to size
 pps – pps systematic
 rppss – randomized pps systematic
 s – systematic
 rg – random group

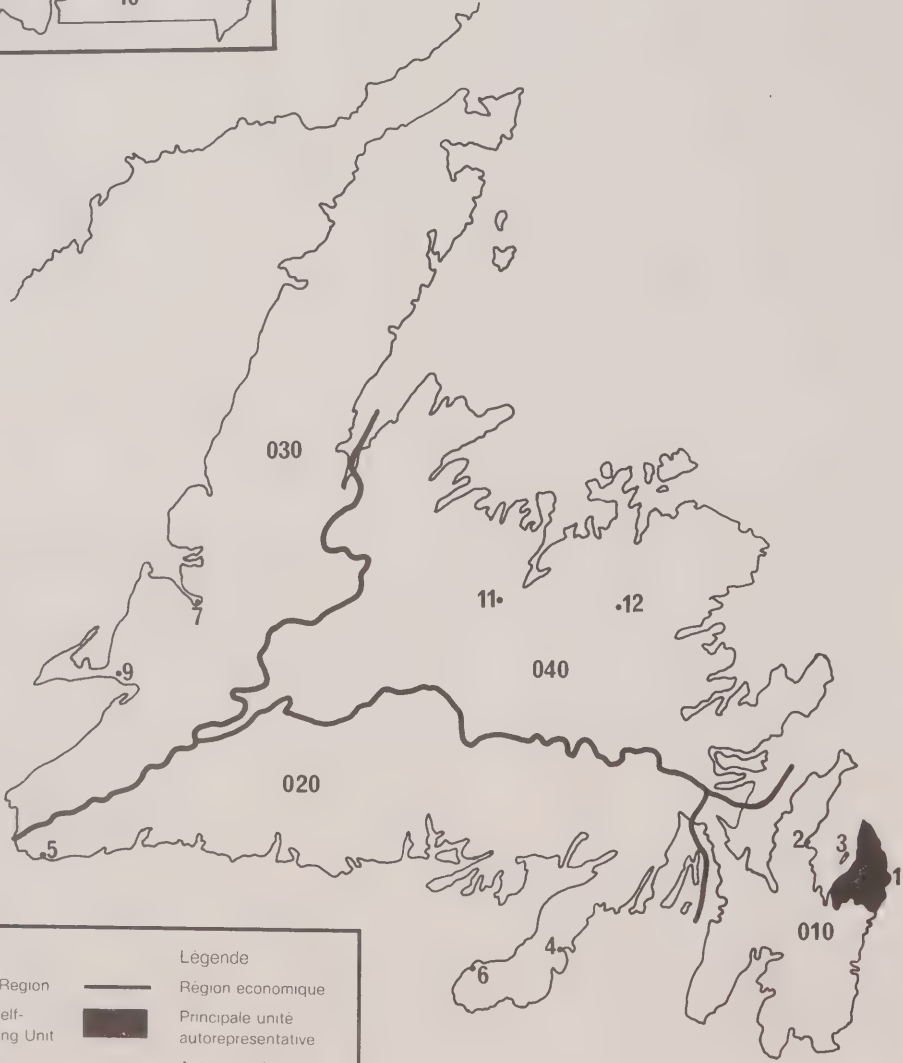
Abréviations :

ppt – probabilité proportionnelle à la taille
 ppts – ppt et tirage systématique
 rppts – ppts avec classement aléatoire
 s – systématique
 ga – groupe aléatoire

Appendix E

Annexe E

NEWFOUNDLAND — TERRE-NEUVE



Legend

Economic Region

Principal Self-Representing Unit

Other Self-Representing Units



Légende

Région économique

Principale unité autoreprésentative

Autres unités autoreprésentatives

Newfoundland

Table 0.2

Non Self-Representing Areas

Economic Region	Stratum	Stratum Type	LF Population ¹	Number of PSUs	Number selected	ISR ²
Région économique	Strate	Type de strate	Population active ¹	Nombre de UPE	Nombre sélectionné	FSI ²
010	01000	urban/urbaine	14068	7	2	49
	01020	rural/rurale	15506	9	2	49
	01040	rural/rurale	18708	9	2	49
020	02000	rural/rurale	9916	7	2	32
	02020	mixed/mixte	11105	7	2	32
030	03000	rural/rurale	14943	7	2	49
	03020	rural/rurale	13849	7	2	49
	03040	rural/rurale	14558	7	2	49
040	04000	urban/urbaine	15504	9	2	49
	04020	urban/urbaine	15050	8	2	49
	04040	rural/rurale	15471	10	2	49
	04060	rural/rurale	15656	10	2	49
	04080	rural/rurale	17600	10	2	49
TOTAL	13	-	191934	107	26	-

¹ Non-institutional civilian population 15 years and over – Personnes civiles non pensionnaires d'institution âgées de 15 ans et plus

² Inverse sampling rate. – Fraction de sondage inverse

ER CD COMPOSITION
RE COMPOSITION en DR

010 Division 1

020 Divisions 2, 3

030 Divisions 4, 5, 9, 10

040 Divisions 6, 7, 8

Terre-Neuve

Tableau 0.2

Unités non autoreprésentatives

Table 0.3

Self-Representing Area List

Tableau 0.3

Liste des unités autoreprésentatives

Serial No. ¹	Name of SRU	1981 LF Population	No. Of Sub-Units	Numerical Identification	Economic Region	Cluster Design ²	Design ISR ³
No. de Série ¹	Nom de l'UAR	Population active de 1981	No de sous-unités	Code d'identification	Région économique	Formation des grappes ²	FSI selon le plan ³
1	St. John's	111810	16	01101	010	BF EA	68(91)
2	Carbonear	7520	2	01201	010	VR	68
3	Wabana-Bell Island	3354	1	01301	010	VR	68
4	Marystown	7415	1	02101	020	VR	32
5	Channel-Port aux Basques	5115	1	02201	020	VR	32
6	Fortune-Grand Bank	4358	1	02301	020	VR	32
7	Corner Brook	18337	2	03101	030	VR	57
8	Labrador-Wabush	9441	2	03201	030	VR	68
9	Stephenville	7664	2	03301	030	VR	68
10	Happy Valley-Goose Bay	4557	1	03401	030	VR	68
11	Windsor-Grand Falls	10349	1	04101	040	VR	68
12	Gander	7216	2	04201	040	VR	68
TOTAL		197136	32				

¹ As on the map. – Voir la carte

² BF = Block-face counts from Area Master File used to form clusters – Utilisant les chiffres pour les côtés d'îlots provenant du fichier principal des régions

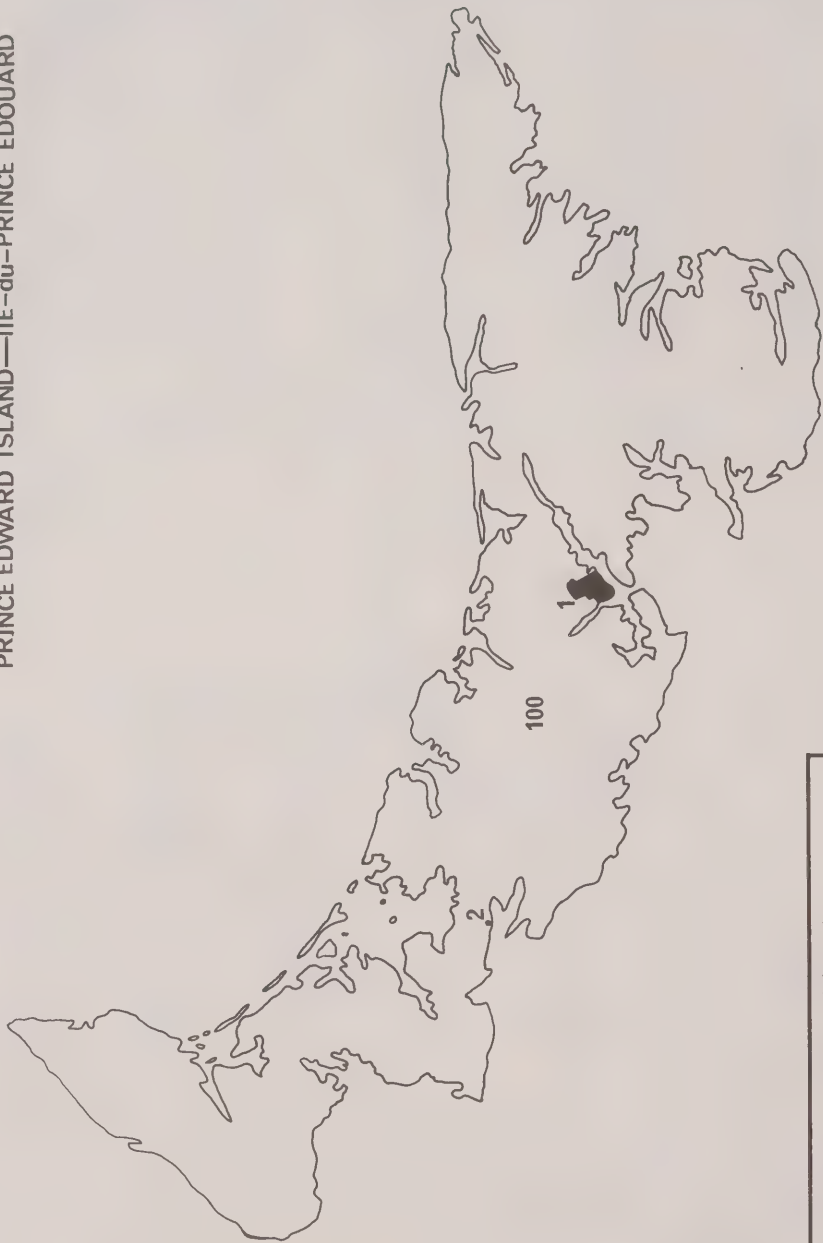
VR = Counts derived from Visitation Records used to form clusters – Utilisant les chiffres provenant des registres des visites

EA = Enumeration Areas used to form clusters – Utilisant les secteurs de dénombrement

³ ISR = Inverse Sampling Rate – (after 1986 sample reduction)

FSI = Fraction de sondage inverse (après la réduction de la taille de l'échantillon en 1986)

PRINCE EDWARD ISLAND—ÎLE-DU-PRINCE ÉDOUARD



Legend

Principal Self-Representing Unit

Other Self-Representing Units

Légende

Principale unité autoreprésentative

Autres unités autoreprésentatives

■

● 2

Prince Edward Island

Table 1.2

Non Self-Representing Areas

Economic Region	Stratum	Stratum Type ³	LF Population ¹	Number of PSUs	Number selected	ISR ²
Région économique	Strate	Type de strate ³	Population active ¹	Nombre de UPE	Nombre sélectionné	FSI ²
100	10000	EA	62381	12	12	25

¹ Non-institutional, civilian population 15 year and over. – Personnes civiles non pensionnaires d'institution âgées de 15 ans et plus.

² Inverse sampling rate. – Fraction de sondage inverse.

³ Special design consisting of Enumeration Areas selected as PSUs. – Plan de sondage special consistant de SD.

Île-du-Prince-Édouard

Tableau 1.2

Unités non autoreprésentatives

Table 1.3

Self-Representing Area List

Tableau 1.3

Liste des unités autoreprésentatives

Serial No. ¹	Name of SRU	1981 LF Population	No. Of Sub-Units	Numerical Identification	Economic Region	Cluster Design ²	Design ISR ³
No. de Série ¹	Nom de l'UAR	Population active de 1981	No. de sous-unités	Code d'identification	Région économique	Formation des grappes ²	FSI selon le plan ³
1	Charlottetown	19361	9	10101	100	VR	30
2	Summerside	8029	4	10201	100	VR	30
TOTAL		27390	13	–	–	–	–

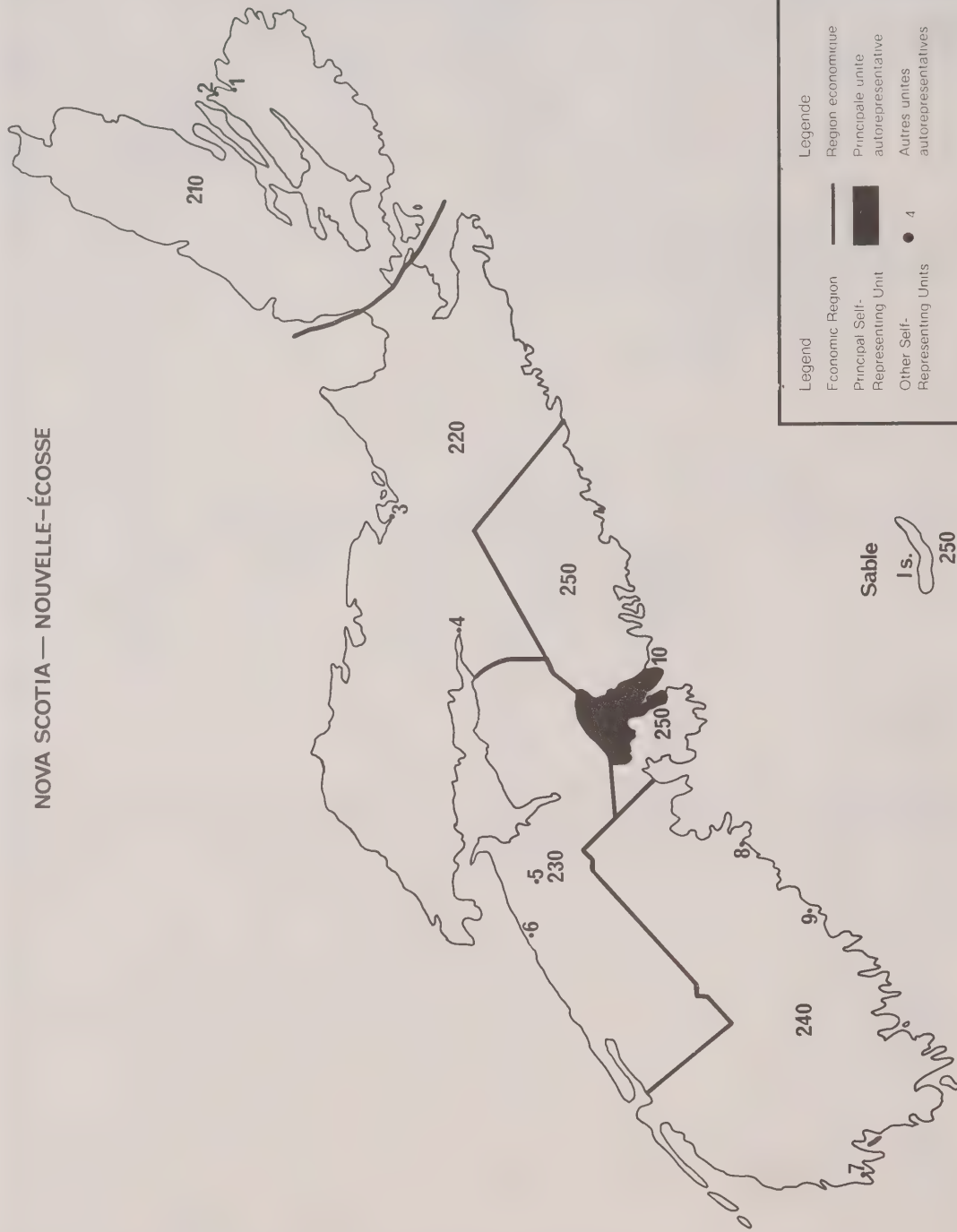
¹ As on the map. – Voir la carte.

² VR = Counts derived from Visitation Records used to form clusters. – Utilisant les chiffres provenant des registres des visites.

³ ISR = Inverse Sampling Rate. – (after 1986 sample reduction)

FSI = Fraction de sondage inverse (après la réduction de la taille de l'échantillon en 1986)

NOVA SCOTIA — NOUVELLE-ÉCOSSE



Legend

— Economic Region

■ Principal Self-Representing Unit

• Other Self-Representing Units

Legende

— Région économique

■ Principale unité auto-représentative

• Autres unités auto-représentatives

Sable

1 s. 250

Nova Scotia

Table 2.2

Non Self-Representing Areas

Nouvelle-Écosse

Tableau 2.2

Unités non autoreprésentatives

Economic Region	Stratum	Stratum Type	LF Population ¹	Number of PSUs	Number selected	ISR ²
Région économique	Strate	Type de strate	Population active ¹	Nombre de UPE	Nombre sélectionné	FSI ²
210	21000	urban/urbaine	33836	17	3	90
220	22000	urban/urbaine	20669	5	2	90
	22020	rural/rurale	19917	13	2	90
	22040	rural/rurale	21842	13	2	90
	22060	rural/rurale	22767	13	2	90
230	23000	urban/urbaine	9923	5	2	50
	23020	rural/rurale	14485	10	2	50
	23040	rural/rurale	15277	10	2	50
	23060	rural/rurale	13150	7	2	50
	23080	rural/rurale	12914	7	2	50
240	24000	urban/urbaine	10306	5	2	50
	24020	rural/rurale	25012	14	3	50
	24040	rural/rurale	22293	14	3	50
	24060	rural/rurale	21149	14	3	50
250	25080	EA dsgn/SD	7951	1	1	90
TOTAL	15	-	271491	148	33	-

¹ Non-institutional civilian population 15 year and over. - Personnes civiles non pensionnaires d'institution âgées de 15 ans et plus

² Inverse sampling rate. - Fraction de sondage inverse

ER CD COMPOSITION

ER COMPOSITION EN DR

210 Divisions 15,16,17,18

220 Divisions 10,11,12,13,14

230 Divisions 05,07,08

240 Divisions 01,02,03,04,06

250 Division 09

Table 2.3

Self-Representing Area List

Tableau 2.3

Liste des unités autoreprésentatives

Serial No. ¹	Name of SRU	1981 LF Population	No Of Sub-Units	Numerical Identification	Economic Region	Cluster Design ²	Design ISR ³
No de Série ¹	Nom de l'UAR	Population active de 1981	No de sous-unités	Code d'identification	Région économique	Formation des grappes ²	FSI selon le plan ³
1	Sydney-Glace Bay	65223	9	21101	210	VR	80 (100)
2	Sydney Mines	24709	3	21201	210	VR	80
3	New Glasgow	17832	3	22101	220	VR	58
4	Truro	14700	2	22201	220	VR	58
5	Kentville	7224	1	23101	230	VR	50
6	Kingston-Greenwood	3526	1	23201	230	VR	50
7	Yarmouth	5769	1	24101	240	VR	50
8	Bridgewater	5259	1	24201	240	VR	50
9	Liverpool	4857	1	24301	240	VR	50
10	Halifax	206024	26	25101	250	BF EA	90 (120)
	Halifax APT		1	25100	250	APT	90 (120)
TOTAL		355123	48	REGULAR			
			1	APT			

¹ As on the map. - Voir la carte.

² BF = Block-face counts from Area Master File used to form clusters - Utilisant les chiffres pour les côtés d'îlots provenant du fichier principal des régions

VR = Counts derived from Visitation Records used to form clusters - Utilisant les chiffres provenant des registres des visites

EA = Enumeration Areas used to form clusters - Utilisant les secteurs de dénombrement

APT = Apartment Buildings used to form clusters - Utilisant les immeubles d'appartements

³ ISR = Inverse Sampling Rate. - (after 1986 sample reduction)

FSI = Fraction de sondage inverse (après la réduction de la taille de l'échantillon en 1986)

NEW BRUNSWICK — NOUVEAU-BRUNSWICK



Legend

Economic Region

Principal Self-Representing Unit

Other Self-Representing Units

Légende

Région économique

Principale unité autoreprésentative

Autres unités autoreprésentatives

New Brunswick

Table 3.2

Non Self-Representing Areas

Economic Region	Stratum	Stratum Type	LF Population ¹	Number of PSUs	Number selected	ISR ²
Région économique	Strate	Type de strate	Population active ¹	Nombre de UPE	Nombre sélectionné	FSI ²
310	31000	urban/urbaine	13488	6	2	60
	31020	rural/rurale	18526	11	2	60
	31040	rural/rurale	18350	11	2	60
	31060	rural/rurale	16633	11	2	60
	31080	rural/rurale	19806	11	2	60
320	32000	urban/urbaine	12401	5	2	60
	32020	rural/rurale	18425	11	2	60
	32040	rural/rurale	15486	11	2	60
330	33000	mixed/mixte	19969	12	2	60
	33020	mixed/mixte	16980	12	2	60
340	34000	rural/rurale	21373	13	2	60
	34020	rural/rurale	18298	13	2	60
350	35000	urban/urbaine	8010	5	2	45
	35020	rural/rurale	10388	9	2	45
	35040	rural/rurale	11254	9	2	45
	35060	rural/rurale	12101	9	2	45
TOTAL	16	-	251488	159	32	-

¹ Non-institutional civilian population 15 year and over - Personnes civiles non pensionnaires d'institution âgées de 15 ans et plus

² Inverse sampling rate - Fraction de sondage inverse

ER CD COMPOSITION
RE COMPOSITION EN DR

310 Divisions 09,14,15

320 Divisions 06,07,08

330 Divisions 01,02,05

340 Divisions 03,04,10

350 Divisions 11,12,13

Nouveau-Brunswick

Tableau 3.2

Unités non autoreprésentatives

Tableau 3.3

Liste des unités autoreprésentatives

Serial No. ¹	Name of SRU	1981 LF Population	No. Of Sub-Units	Numerical Identification	Economic Region	Cluster Design ²	Design ISR ³
No de Série ¹	Nom de l'UAR	Population active de 1981	No de sous-unités	Code d'identification	Région économique	Formation des grappes ²	FSI selon le plan ³
1	Dalhousie-Campbellton	14324	4	31301	310	VR	60
2	Bathurst	14243	4	31101	310	VR	45
3	Chatham/Newcastle	13125	4	31201	310	VR	45
4	Moncton	74448	13	32101	320	BF EA	60(80)
5	Saint John	84917	18	33101	330	BF EA	60(80)
6	Fredericton	33820	11	34101	340	BF EA	45(56)
7	Oromocto	4273	2	34201	340	VR	60
8	Edmundston	13400	4	35101	350	VR	45
9	Grand Falls	4477	1	35201	350	VR	45
TOTAL		257027	61				

¹ As on the map. - Voir la carte.

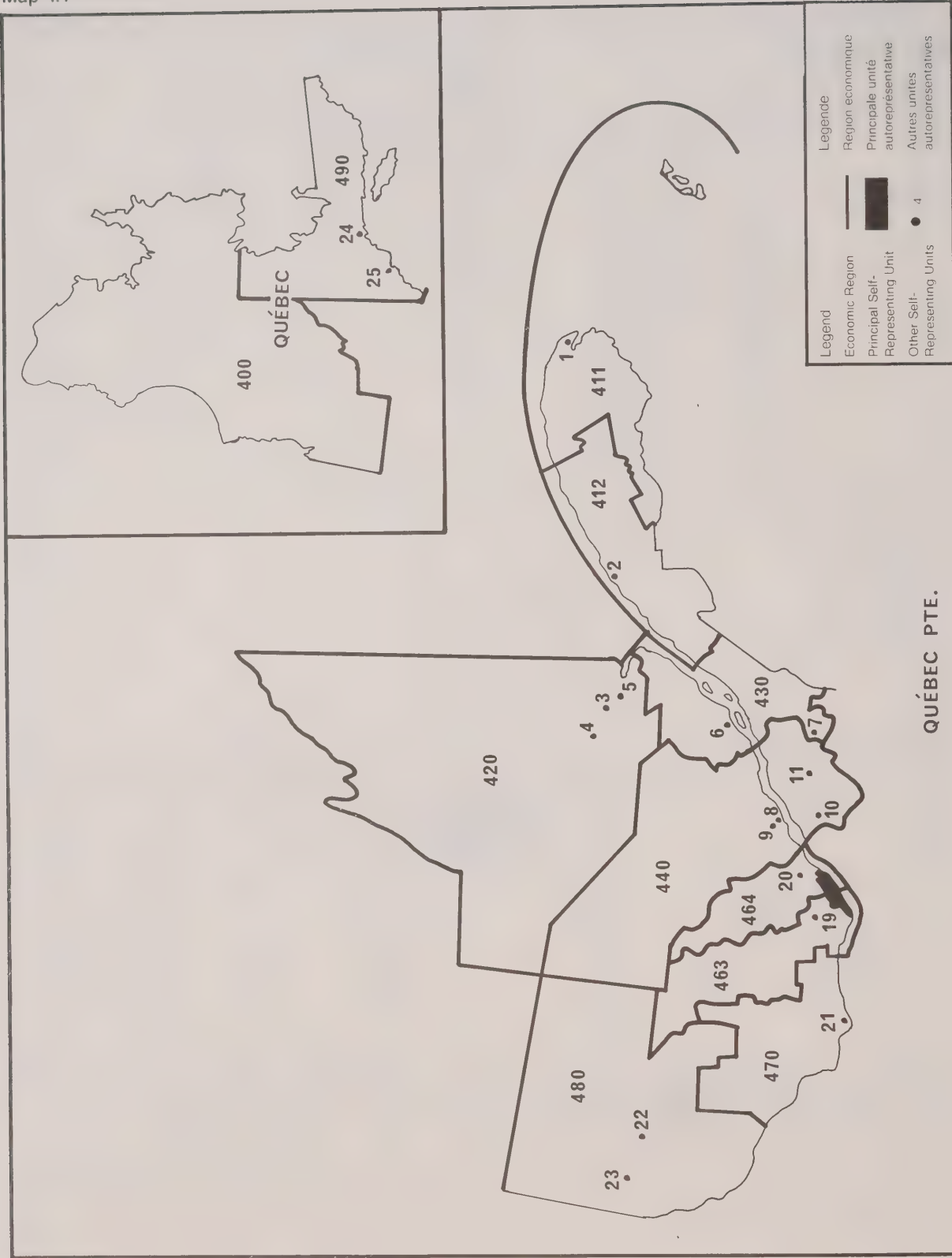
² BF = Block-face counts from Area Master File used to form clusters - Utilisant les chiffres pour les côtes d'îlots provenant du fichier principal des régions

VR = Counts derived from Visitation Records used to form clusters - Utilisant les chiffres provenant des registres des visites

EA = Enumeration Areas used to form clusters - Utilisant les secteurs de dénombrement

³ ISR = Inverse Sampling Rate. - (after 1986 sample reduction)

FSI = Fraction de sondage inverse (après la réduction de la taille de l'échantillon en 1986)



QUÉBEC PTE.

Quebec

Table 4.2

Non Self-Representing Areas

Economic Region	Stratum	Stratum Type	LF Population ¹	Number of PSUs	Number selected	ISR ²
Région économique	Strate	Type de strate	Population active ¹	Nombre de UPE	Nombre sélectionné	FSI ²
411	41000	mixed/mixte	71876	13	3	150
412	41020	urban/urbaine	55545	12	2	220
	41040	rural/rurale	79469	17	2	220
420	42000	mixed/mixte	73168	16	2	245
430	43000	mixed/mixte	86145	16	2	245
	43020	mixed/mixte	80858	16	2	245
	43040	mixed/mixte	72587	14	2	245
440	44000	mixed/mixte	68794	16	2	245
	44020	mixed/mixte	84045	16	2	245
450	45000	urban/urbaine	48975	12	2	245
	45020	rural/rurale	54821	16	2	245
461	46000	mixed/mixte	95537	18	3	245
	46020	mixed/mixte	105788	18	3	245
463	46040	mixed/mixte	108955	18	3	245
464	46060	mixed/mixte	48516	12	2	150
	46080	mixed/mixte	41822	14	2	150
470	47000	mixed/mixte	54485	16	2	245
480	48000	mixed/mixte	61648	16	3	150
TOTAL	18	-	1293034	276	41	-

¹ Non-institutional civilian population 15 year and over – Personnes civiles non pensionnaires d'institution âgées de 15 ans et plus

² Inverse sampling rate. – Fraction de sondage inverse

ER CD COMPOSITION
RE COMPOSITION EN DR

400 Divisions 98PT (REMOTE/ ELOIGNEES)
411 Divisions 01,02,03,04PT,06PT
412 Divisions 04PT,05,06PT,07,08,09,10
420 Divisions 17PT,90,93,94,98PT
430 Divisions 11,12,13,14,15,16,17PT,20PT,21,22,23,24PT, 26PT,27PT,28PT,29PT,97PT
440 Divisions 20PT,26PT,27PT,28PT,29PT,32,33,34PT,41,42PT, 43PT,47PT,84PT
450 Divisions 24PT,25,26PT,34PT,35,36,37,38PT,39PT
461 Divisions 38PT,39PT,40,42PT,50,51,52,53,54,55,56,57, 66,67,68,69,70,71,72
462 Divisions 64,65
463 Divisions 47PT,49PT,58PT,61PT,63PT,73,74,75PT,76,78PT
464 Divisions 43PT,47PT,49PT,58PT,61PT,62,63PT
470 Divisions 75PT,78PT,79,80PT
480 Divisions 80PT,83,84PT
490 Divisions 97PT,98PT (REMOTE and SRU ONLY éloignées et UAR seulement)

Quebec

Table 4.3

Self-Representing Area List

Québec

Tableau 4.3

Liste des unités autoreprésentatives

Serial No ¹	Name of SRU	1981 LF Population	No. Of Sub-Units	Numerical Identification	Economic Region	Cluster Design ²	Design ISR ³
No de Série ¹	Nom de l'UAR	Population active de 1981	No. de sous- unités	Code d'identifi- cation	Région écono- mique	Formation des grappes ²	FSI selon le plan ³
1	Gaspé	12830	1	41201	411	VR	40
2	Rimouski	24536	2	41101	412	VR	84
3	Chicoutimi-Jonquière	101125	11	42101	420	BF/EA	128
4	Alma	19371	1	42201	420	VR	286
5	La Baie	14656	1	42301	420	VR	286
6	Ville de Québec	438442	19	43101	430	BF/EA	320 (356)
	" APT		1	43100	430	APT	320 (356)
7	Thetford Mines	23172	1	43201	430	VR	286
8	Trois Rivières	87284	8	44101	440	BF/EA	128
9	Shawinigan	43479	2	44201	440	VR	286
10	Drummondville	34451	2	44301	440	VR	286
11	Victoriaville	25242	1	44401	440	VR	286
12	Sherbrooke	89365	8	45101	450	BF/EA	128
13	Montréal	2224374	51	46101	462	BF/EA	444 (666)
	"		12	46200	461	BF/EA	444 (666)
	"		10	46220	462	BF EA	286 (429)
	"		5	46240	463	BF/EA	286 (429)
	"		4	46250	464	BF EA	286 (429)
	" APT		4	46100	462	APT	444 (666)
	" APT		1	46200	461	APT	444 (666)
	" APT		1	46220	462	APT	286 (429)
14	St-Jean	39574	2	46301	461	VR	286
15	Granby	34319	2	46601	461	VR	286
16	Sorel	33337	2	46501	461	VR	286
17	Saint-Hyacinthe	31707	2	46401	461	VR	286
18	Salaberry-De-Valleyfield	30897	2	46801	461	VR	286
19	St-Jérôme	29167	2	46701	463	BF/EA	286
20	Joliette	24031	1	46811	464	VR	286
21	Hull	128079	10	47101	470	BF/EA	186 (248)
	" APT		1	47100	470	APT	186 (248)
22	Val D'Or	21132	1	48201	480	VR	150
23	Rouyn-Noranda	19947	1	48101	480	VR	150
24	Sept Îles	29533	2	49101	490	VR	105
25	Baie-Comeau	24735	2	49201	490	VR	84
TOTAL		3584785	168		REGULAR		
			8		APT		

¹ As on the map. - Voir la carte.² BF = Block-face counts from Area Master File used to form clusters. - Utilisant les chiffres pour les côtés d'îlots provenant du fichier principal des régions.

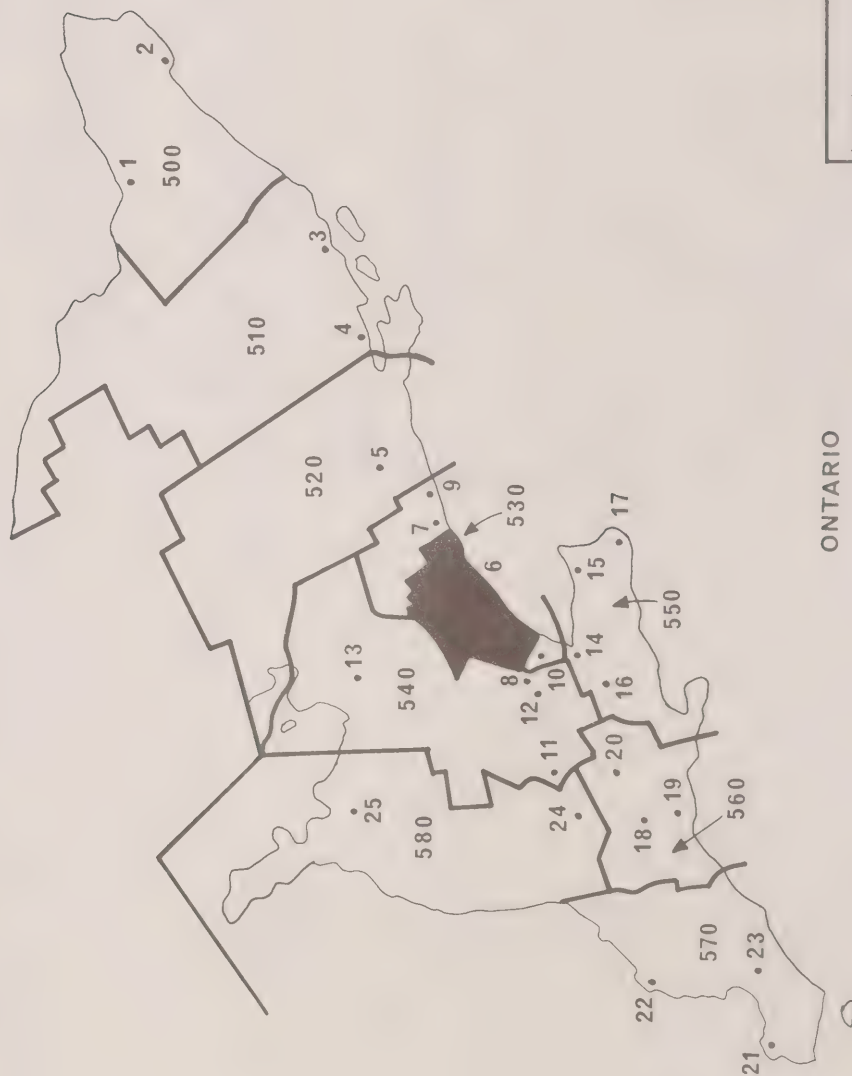
VR = Counts derived from Visitation Records used to form clusters. - Utilisant les chiffres provenant des registres des visites.

EA = Enumeration Areas used to form clusters. - Utilisant les secteurs de dénombrement.

APT = Apartment Buildings used to form clusters. - Utilisant les immeubles d'appartements.







³ ISR = Inverse Sampling Rate. - (after 1986 sample reduction).

FSI = Fraction de sondage inverse (après la réduction de la taille de l'échantillon en 1986).

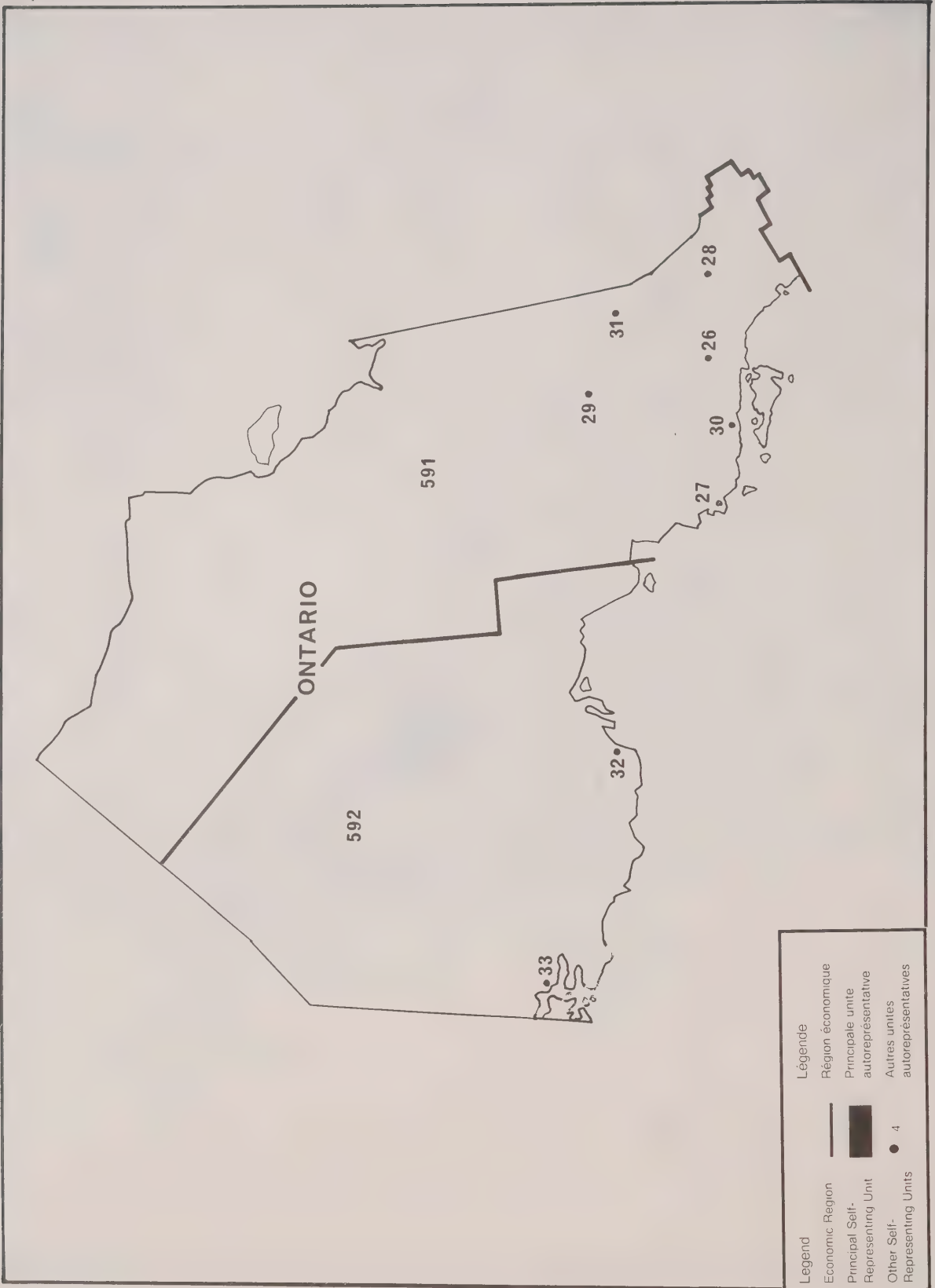


ONTARIO

Legend

	Unit boundaries		Unit boundaries
	Principal Self-Representing Unit		Principale unite autorepresentative
	Population centre		Centre de population

Source: Statistics Canada, 1991



Ontario

Table 5.2

Non Self-Representing Areas

Economic Region	Stratum	Stratum Type	LF Population ¹	Number of PSUs	Number selected	ISR ²
Région économique	Strate	Type de strate	Population active ¹	Nombre de UPE	Nombre sélectionné	FSI ²
500	50000	urban/urbaine	68323	14	2	286
	50020	rural/rurale	107888	19	3	286
510	51000	urban/urbaine	53308	13	2	286
	51020	rural/rurale	96723	19	3	286
520	52000	urban/urbaine	52946	11	2	190
	52020	rural/rurale	45217	14	2	190
	52040	rural/rurale	53799	14	2	190
530	53000	urban/urbaine	18154	7	2	286
	53080	EA dsgn:SD	22250	1	1	286
540	54000	mixed:mixte	64649	15	2	286
	54020	mixed:mixte	65901	16	2	286
	54040	mixed:mixte	69216	16	2	286
550	55000	mixed:mixte	94286	17	2	286
560	56000	mixed:mixte	84601	16	2	286
570	57000	mixed:mixte	114303	18	3	286
	57080	EA dsgn:SD	11760	1	1	286
580	58000	mixed:mixte	41860	12	2	150
	58020	mixed:mixte	38869	12	2	150
	58040	mixed:mixte	36515	12	2	150
	58060	mixed:mixte	39170	12	2	150
591	59000	mixed:mixte	64725	16	2	278
	59020	mixed:mixte	85718	16	2	278
592	59040	urban/urbaine	46382	14	2	190
TOTAL	23	-	1376563	305	47	-

¹ Non-institutional civilian population 15 years and over. - Personnes civiles non pensionnaires d'institution âgées de 15 ans et plus

² Inverse sampling rate. - Fraction de sondage inverse

ER CD COMPOSITION
RE COMPOSITION EN DR

500 Divisions 01,02,03,04,05,06,07,08,09
510 Divisions 10,11,12,13,47
520 Divisions 14,15,16,44,46
530 Divisions 18,19,20,21,24PT
540 Divisions 22,23,30,43
550 Divisions 24PT,25,26,28,29
560 Divisions 32,34,39
570 Divisions 36,37,38
580 Divisions 31,40,41,42
591 Divisions 48,49,51,52,53,54,56,57,60PT
592 Divisions 58,59,60PT

Ontario

Table 5.3

Self-Representing Area List

Ontario

Tableau 5.3

Liste des unités autoreprésentatives

Serial No. ¹	Name of SRU	1981 LF Population	No. Of Sub-Units	Numerical Identification	Economic Region	Cluster Design ²	Design ISR ³
No. de Série ¹	Nom de l'UAR	Population active de 1981	No. de sous-unités	Code d'identification	Région économique	Formation des grappes ²	FSI selon le plan ³
1	Ottawa	424149	17	50101	500	BF/EA	340 (454)
	" APT		2	50100	500	APT	340 (454)
2	Cornwall	35727	2	50201	500	VR	130
3	Kingston	71497	4	51101	510	BF/EA	286
4	Belleville-Trenton	43134	3	51201	510	BF/EA	286
5	Peterborough	51848	4	52101	520	VR	190
6	Toronto	2356595	62	53101	530	BF/EA	400 (600)
	Toronto			53601	530	BF/EA	400 (600)
	" APT		12	53100	530	APT	400 (600)
	" APT			53600	530	APT	400 (600)
7	Oshawa	115105	8	53201	530	BF/EA	170
	" APT		1	53200	530	APT	170
8	Halton Hills	26263	1	53301	530	VR	307
9	Newcastle	23940	1	53401	530	VR	307
10	Milton	19905	1	53501	530	VR	307
11	Kitchener-Waterloo	217326	11	54101	540	BF/EA	307
	" APT		1	54100	540	APT	307
12	Guelph	56735	5	54201	540	BF/EA	190
13	Barrie	41943	2	54301	540	VR	307
14	Hamilton-Burlington	422357	15	55101	550	BF/EA	340 (378)
	" APT		2	55100	550	APT	340 (378)
15	St. Catharines-Niagara	236366	10	55201	550	BF/EA	307
	" APT		1	55200	550	APT	307
16	Brantford	67038	4	55301	550	VR	246
17	Fort Erie	18748	1	55401	550	VR	307
18	London	219546	11	56101	560	BF/EA	307
	" APT		1	56100	560	APT	307
19	St. Thomas	28618	1	56201	560	VR	307
20	Woodstock	26909	1	56301	560	VR	307
21	Windsor	187955	9	57101	570	BF/EA	286
	" APT		1	57100	570	APT	286
22	Sarnia	63615	4	57201	570	BF/EA	246
23	Chatham	30287	2	57301	570	VR	307
24	Stratford	20438	2	58101	580	VR	150
25	Owen Sound	15245	1	58201	580	VR	150
26	Sudbury	111058	9	59101	591	BF/EA	150
27	Sault Ste. Marie	62859	4	59201	591	BF/EA	190
28	North Bay	39866	4	59301	591	BF/EA	150
29	Timmins	32816	2	59401	591	VR	307
30	Elliot Lake	11032	1	59501	591	VR	150
31	Kirkland Lake	8957	1	59601	591	VR	150
32	Thunder Bay	92977	11	59701	592	BF/EA	130
33	Kenora-Keewatin	11664	1	59801	592	VR	130
TOTAL		5192518	215			REGULAR	
			21			APT	

¹ As on the map. - Voir la carte.² BF = Block-face counts from Area Master File used to form clusters. - Utilisant les chiffres pour les côtés d'îlots provenant du fichier principal des régions.

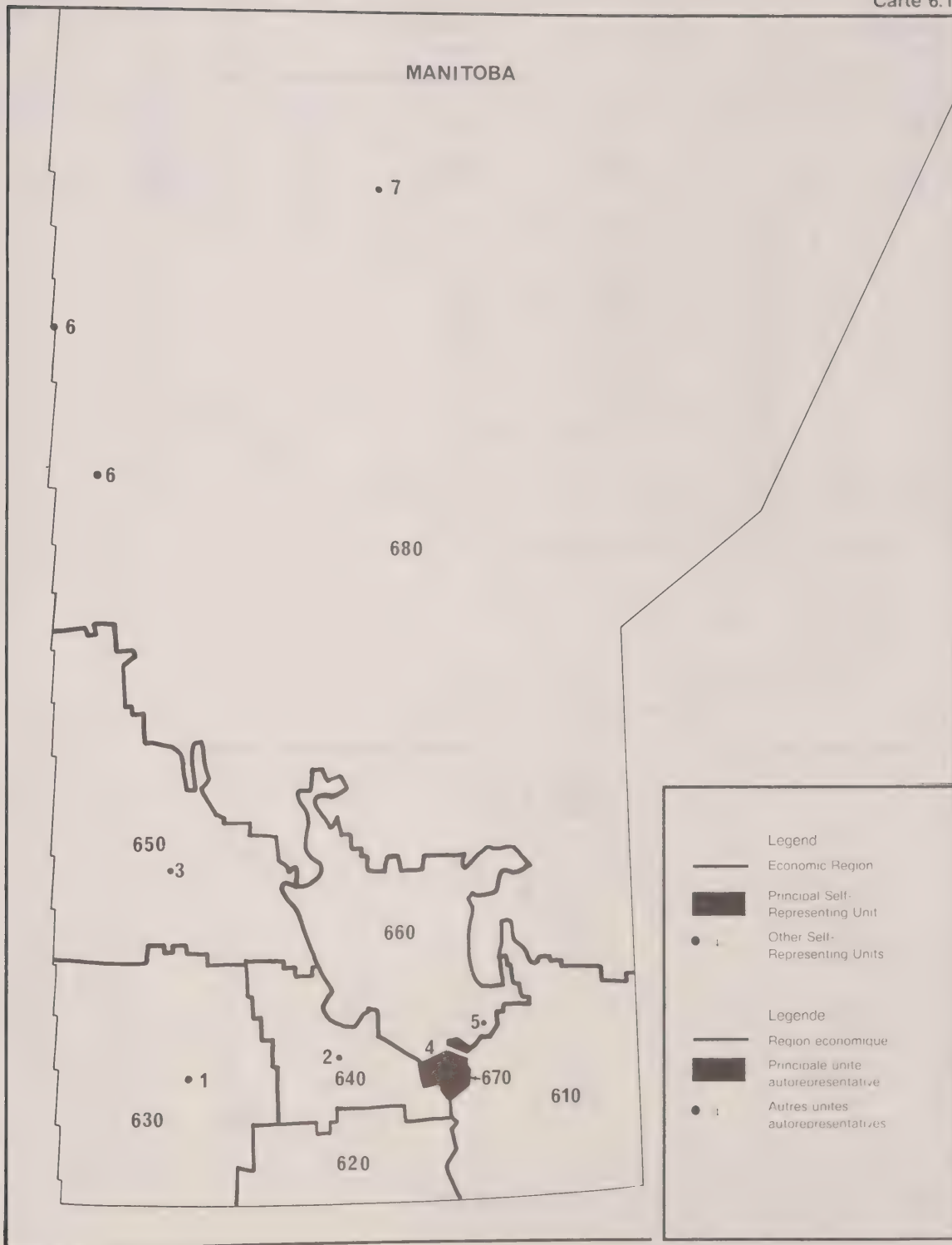
VR = Counts derived from Visitation Records used to form clusters. - Utilisant les chiffres provenant des registres des visites.

EA = Enumeration Areas used to form clusters. - Utilisant les secteurs de dénombrement.

APT = Apartment Buildings used to form clusters. - Utilisant les immeubles d'appartements.

³ ISR = Inverse Sampling Rate. - (after 1986 sample reduction).

FSI = Fraction de sondage inverse (après la réduction de la taille de l'échantillon en 1986)



Manitoba

Table 6.2

Non Self-Representing Areas

Economic Region	Stratum	Stratum Type	LF Population ¹	Number of PSUs	Number selected	ISR ²
Région économique	Strate	Type de strate	Population active ¹	Nombre de UPE	Nombre sélectionné	FSI ²
610	61000	urban/urbaine	9996	5	2	55
	61020	rural/rurale	13380	10	2	55
	61040	rural/rurale	16172	10	2	55
620	62000	urban/urbaine	12307	5	3	40
	62020	rural/rurale	11807	8	2	40
	62040	rural/rurale	11025	8	2	40
630	63000	urban/urbaine	16207	11	2	65
	63020	rural/rurale	17400	12	2	65
	63040	rural/rurale	18558	12	2	65
640	64000	rural/rurale	11551	8	2	40
	64020	rural/rurale	10360	8	2	40
	65000	rural/rurale	15831	10	2	65
650	65020	mixed/mixte	15819	10	2	65
	66000	rural/rurale	17637	10	2	65
660	66020	rural/rurale	18411	10	2	65
TOTAL	15	-	216461	137	31	

¹ Non-institutional civilian population 15 years and over. - Personnes civiles non pensionnaires d'institution âgées de 15 ans et plus.

² Inverse sampling rate - Fraction de sondage inverse

ER CD COMPOSITION
RE COMPOSITION EN DR

610 Divisions 01,02,12
620 Divisions 03,04
630 Divisions 05,06,07,15
640 Divisions 08,09,10
650 Divisions 16,17,20
660 Divisions 13,14,18
670 Division 11 (SRU ONLY/UAR seulement)
680 Divisions 19,21,22,23 (REMOTE and SRU ONLY / éloignées et UAR seulement)

Table 6.3

Self-Representing Areas List

Tableau 6.3

Liste des unités autoreprésentatives

Serial No. ¹	Name Of SRU	1981 LF Population	No. Of Sub-Units	Numerical Identification	Economic Region	Cluster Design ²	Design ISR ³
No. de Série ¹	Nom de l'UAR	Population active de 1981	No. de sous-unités	Code d'identification	Région économique	Formation des grappes ²	FSI selon le plan ³
1	Brandon	27257	5	63101	630	VR	80
2	Portage La Prairie	9257	2	64101	640	VR	40
3	Dauphin	7140	1	65101	650	VR	80
4	Winnipeg	453441	1	61101	610	EA	117 (156)
	"		1	66101	660	EA	117 (156)
	"		53	67101	670	BF/EA	117 (156)
	"			67201	670	BF/EA	117 (156)
	" APT		2	67100	670	APT	117 (156)
	" APT			67200	670	APT	117 (156)
5	Selkirk	7161	1	66201	660	VR	80
6	Flin Flon/The Pas	10392	2	68201	680	VR	80
7	Thompson	9486	2	68101	680	VR	80
	TOTAL	524134	68				
			2		REGULAR APT		

¹ As on the map. - Voir la carte

² BF = Block-face counts from Area Master File used to form clusters. - Utilisant les chiffres pour les côtés d'îlots provenant du fichier principal des régions.

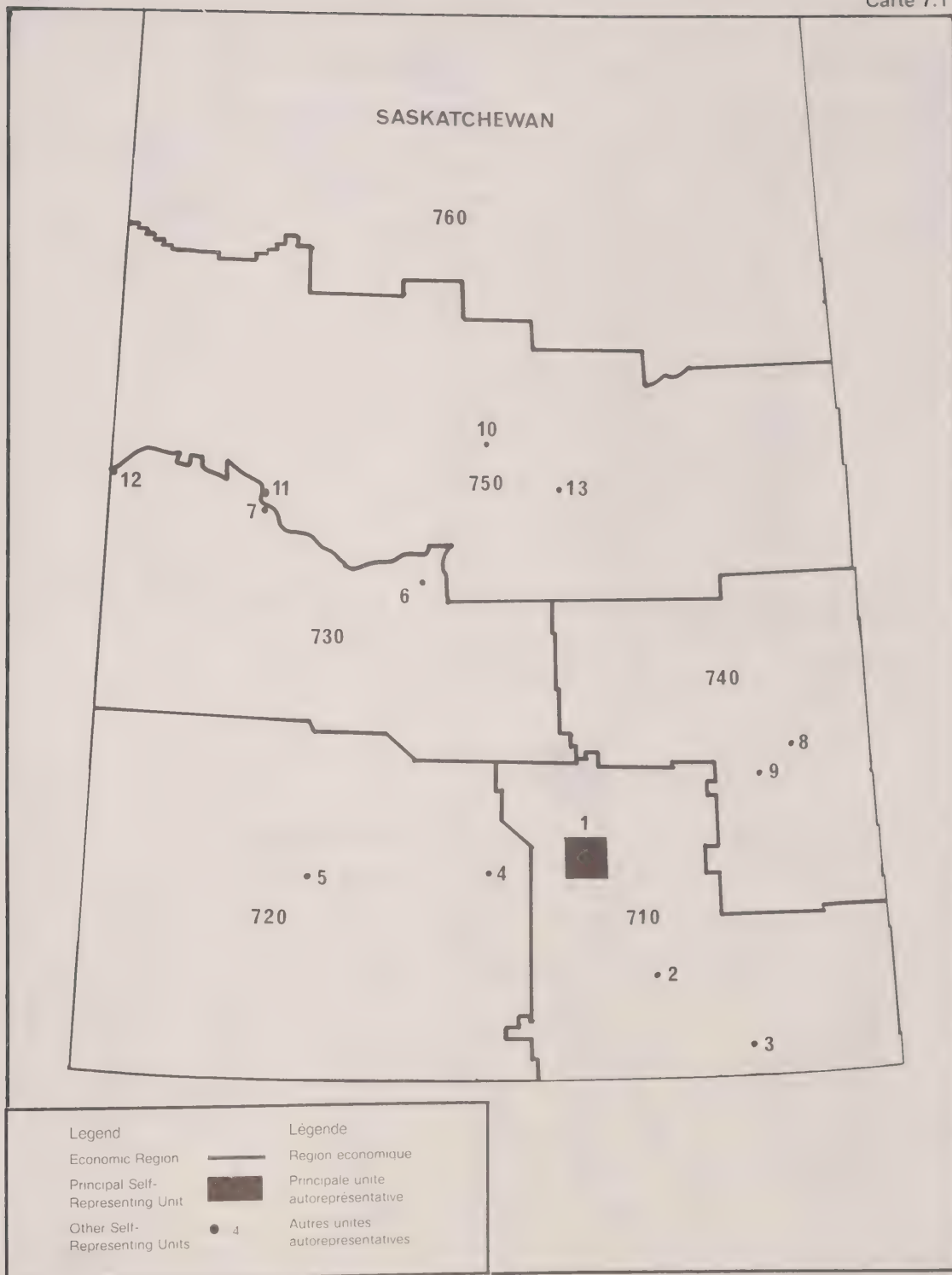
VR = Counts derived from Visitation Records used to form clusters. - Utilisant les chiffres provenant des registres des visites.

EA = Enumeration Areas used to form clusters. - Utilisant les secteurs de dénombrement.

APT = Apartment Buildings used to form clusters. - Utilisant les immeubles d'appartements.

³ FSI = Inverse Sampling Rate (after 1986 sample reduction).

FSI = Fraction de sondage inverse (après la réduction de la taille de l'échantillon en 1986).



Saskatchewan

Table 7.2

Non Self-Representing Areas

Saskatchewan

Tableau 7.2

Unités non autoreprésentatives

Economic Region	Stratum	Stratum Type	LF Population ¹	Number of PSUs	Number selected	ISR ²
Région économique	Strate	Type de strate	Population active ¹	Nombre de UPE	Nombre sélectionné	FSI ²
710	71000	rural/rurale	18929	13	2	75
	71020	rural/rurale	18782	13	2	75
	71040	rural/rurale	17901	13	2	75
720	72000	rural/rurale	18799	13	2	75
	72020	rural/rurale	18821	13	2	75
	72040	mixed/mixte	17181	13	2	75
730	73000	urban/urbaine	17373	11	2	75
	73020	rural/rurale	22480	13	2	75
	73040	rural/rurale	18218	13	2	75
	73080	EA dsgrn/SD	4009	1	1	75
740	74000	urban/urbaine	16424	11	3	45
	74020	rural/rurale	13423	9	2	45
	74040	rural/rurale	12067	9	2	45
	74060	rural/rurale	12407	9	2	45
	74080	rural/rurale	11522	9	2	45
750	75000	urban/urbaine	18571	10	2	75
	75020	rural/rurale	18927	13	2	75
	75040	rural/rurale	21654	13	2	75
	75060	rural/rurale	18339	13	2	75
	75080	rural/rurale	20620	13	2	75
TOTAL	20	-	336447	225	40	-

¹ Non-institutional civilian population 15 years and over. - Personnes civiles non pensionnaires d'institution âgées de 15 ans et plus.

² Inverse sampling rate. - Fraction de sondage inverse.

ER CD COMPOSITION
RE COMPOSITION EN DR

710 Divisions 01,02,06
720 Divisions 03,04,07,08
730 Divisions 11,12,13
740 Divisions 05,09,10
750 Divisions 14,15,16,17
760 Division 18 (remote only / éloignées seulement)

Table 7.3

Self-Representing Area List

Tableau 7.3

Liste des unités autoreprésentatives

Serial No. ¹	Name Of SRU	1981 LF Population	No. Of Sub-Units	Numerical Identification	Economic Region	Cluster Design ²	Design ISR ³
No. de Série ¹	Nom de l'UAR	Population active de 1981	No. de sous-unités	Code d'identification	Région économique	Formation des grappes ²	FSI selon le plan ³
1	Regina	123264	30	71101	710	BF/EA	61(81)
2	Weyburn	7147	2	71201	710	VR	45
3	Estevan	6942	2	71301	710	VR	45
4	Moose Jaw	25351	8	72101	720	VR	45(56)
5	Swift Current	11354	3	72201	720	VR	45
6	Saskatoon	117953	30	73101	730	BF/EA	61(81)
	"APT		1	73100	730	APT	61(81)
7	Battleford	2450	1	73201	730	VR	45
8	Yorkton	11750	3	74101	740	VR	45
9	Melville	3968	1	74201	740	VR	45
10	Prince Albert	22507	6	75101	750	VR	45
11	North Battleford	10606	3	75201	750	VR	45
12	Lloydminster	4431	1	75401	750	VR	45
13	Melfort	4417	1	75301	750	VR	45
	TOTAL	352140	91	REGULAR			
			1	APT			

¹ As on the map. - Voir la carte

² BF = Block-face counts from Area Master File used to form clusters. - Utilisant les chiffres pour les côtés d'îlots provenant du fichier principal des régions.

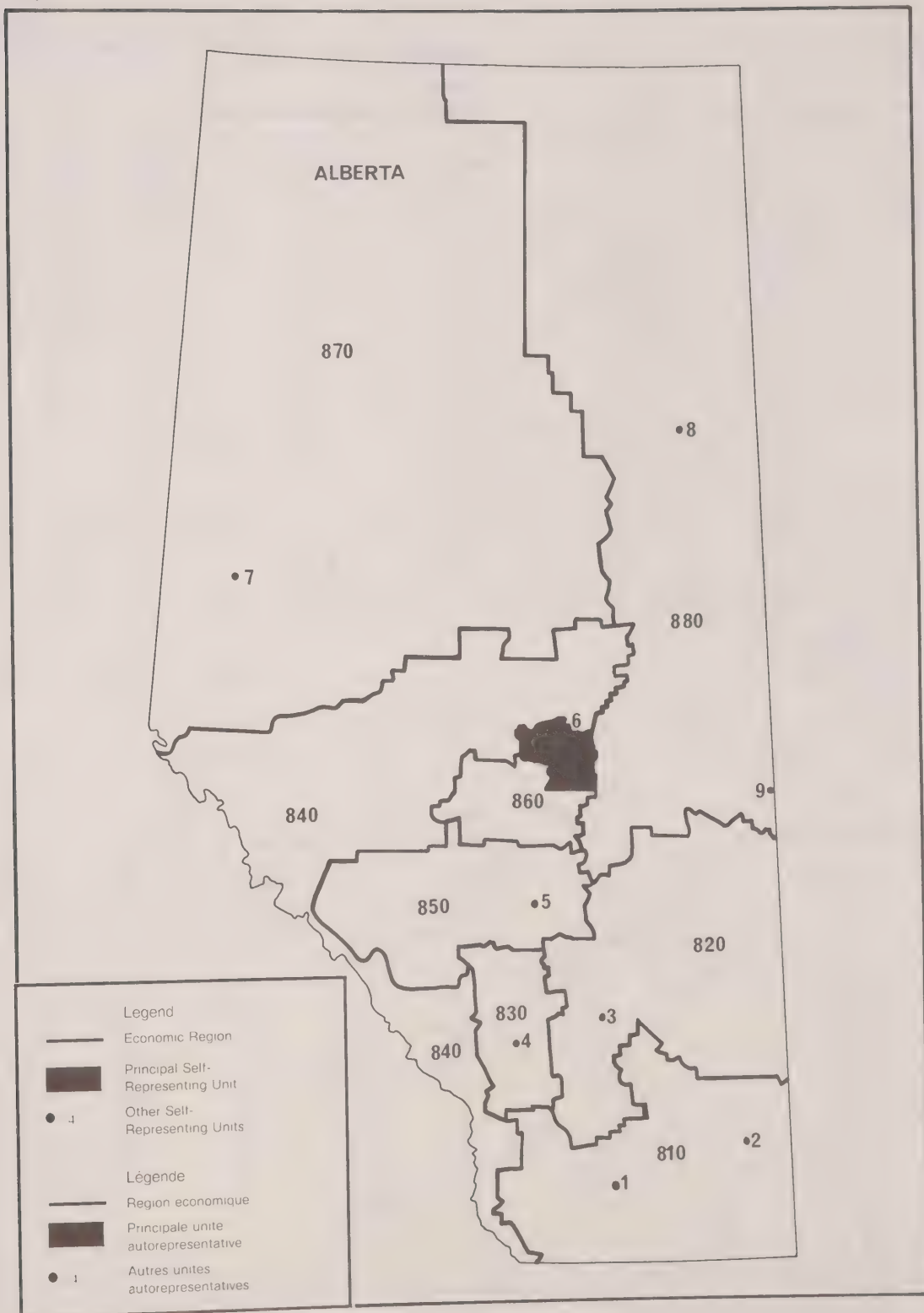
VR = Counts derived from Visitation Records used to form clusters. - Utilisant les chiffres provenant des registres des visites.

EA = Enumeration Areas used to form clusters. - Utilisant les secteurs de dénombrement

APT = Apartment Buildings used to form clusters. - Utilisant les immeubles d'appartements.

³ ISR = Inverse Sampling Rate (after 1986 sample reduction).

FSI = Fraction de sondage inverse (après la réduction de la taille de l'échantillon en 1986).



Alberta

Table 8.2

Non Self-Representing Areas

Economic Region	Stratum	Stratum Type	LF Population ¹	Number of PSUs	Number selected	ISR ²
Région économique	Strate	Type de strate	Population active ¹	Nombre de UPE	Nombre sélectionné	FSI ²
810	81000	urban/urbaine	30280	12	2	160
	81020	rural/rurale	38546	16	2	160
820	82000	urban/urbaine	16586	9	3	45
	82020	rural/rurale	23324	11	4	45
	82040	rural/rurale	19441	11	4	45
830	83000	mixed/mixte	45834	16	2	160
	83080	EA dsgn/SD	8099	1	1	160
840	84000	urban/urbaine	18656	6	3	45
	84020	urban/urbaine	16124	6	3	45
	84040	rural/rurale	17038	9	3	45
	84060	rural/rurale	20484	9	3	45
850	85000	urban/urbaine	20495	6	2	100
	85020	rural/rurale	33272	15	3	100
860	86000	urban/urbaine	31759	6	2	160
	86020	rural/rurale	35920	16	2	160
	86080	EA dsgn/SD	4315	1	1	160
870	87000	urban/urbaine	23694	12	3	60
	87020	rural/rurale	24091	10	3	60
880	88000	urban/urbaine	30997	10	3	100
	88020	rural/rurale	45079	16	3	100
TOTAL	20	-	504034	198	52	-

¹ Non-institutional civilian population 15 years and over. - Personnes civiles non pensionnaires d'institution âgées de 15 ans et plus.

² Inverse sampling rate. - Fraction de sondage inverse

ER CD COMPOSITION
RE COMPOSITION EN DR

810 Divisions 01,02,03
820 Divisions 04,05,07
830 Division 06
840 Divisions 09,13,14
850 Division 08
860 Division 11
870 Division 15
880 Divisions 10,12

Alberta

Tableau 8.2

Unités non autoreprésentatives

Table 8.3

Self-Representing Area List

Tableau 8.3

Liste des unités autoreprésentatives

Serial No. ¹	Name Of SRU	1981 LF Population	No. Of Sub-Units	Numerical Identification	Economic Region	Cluster Design ²	Design ISR ³
No. de Série ¹	Nom de l'UAR	Population active de 1981	No. de sous-unités	Code d'identification	Région économique	Formation des grappes ²	FSI selon le plan ³
1	Lethbridge	41580	7	81101	810	VR	50
2	Medicine Hat	33294	6	81201	810	VR	50
3	Drumheller	4620	1	82101	820	VR	40
4	Calgary	457764	47	83101	830	BF/EA	142 (184)
	" APT		2	83100	830	APT	142 (184)
5	Red Deer	33890	6	85101	850	VR	50
6	Edmonton	501189	51	86101	860	BF/EA	142 (184)
	Edmonton			86201	860	BF/EA	142 (184)
	" APT		2	86100	860	APT	142 (184)
	" APT			86200	860	APT	142 (184)
7	Grande Prairie	17650	3	87101	870	VR	35
8	Fort McMurray	21286	3	88101	880	VR	35
9	Lloydminster	6733	1	88201	880	VR	100
	TOTAL	1118006	125	REGULAR			
		4	APT				

¹ As on the map. - Voir la carte

² BF = Block-face counts from Area Master File used to form clusters. - Utilisant les chiffres pour les côtés d'îlots provenant du fichier principal des régions.

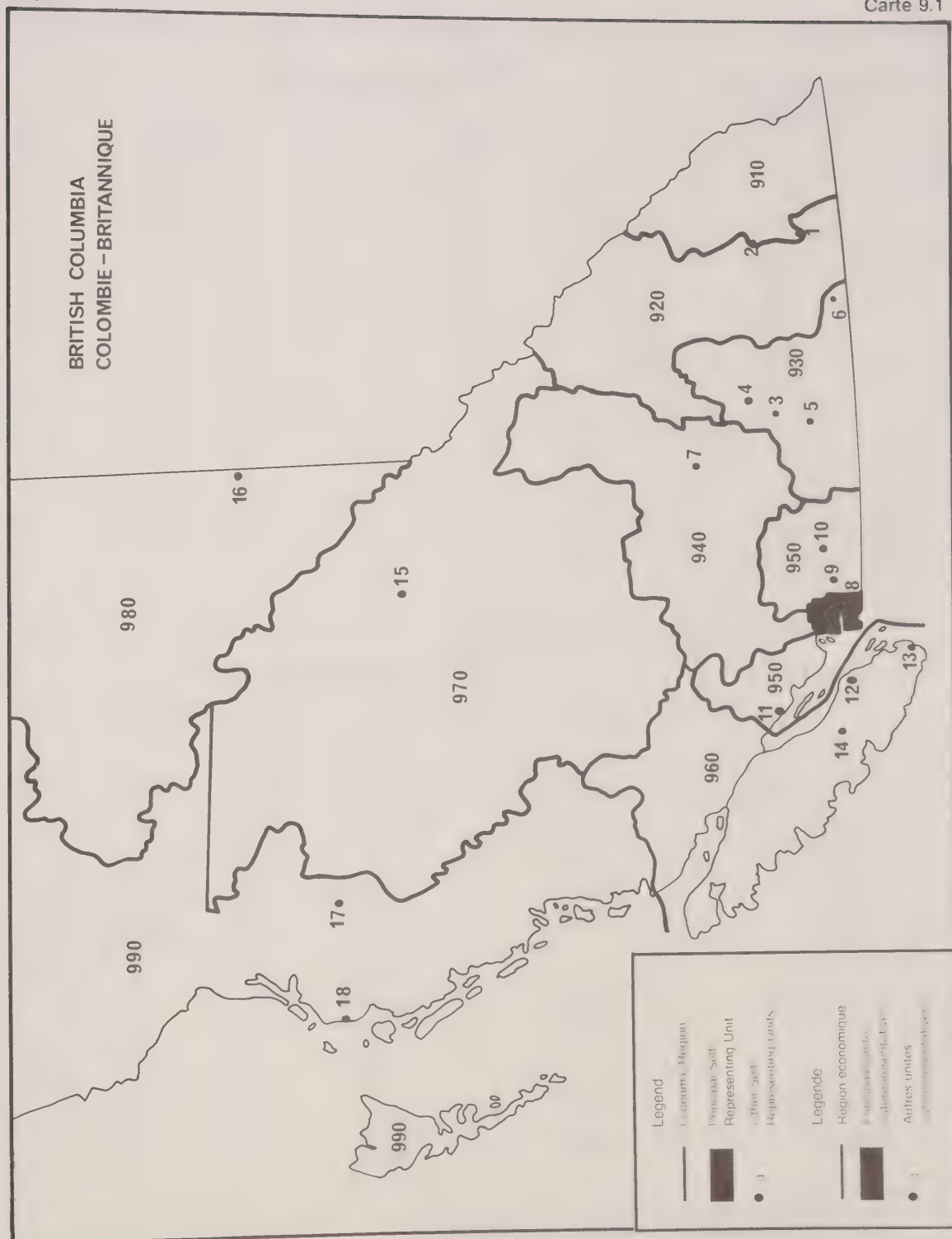
VR = Counts derived from Visitation Records used to form clusters. - Utilisant les chiffres provenant des registres des visites.

EA = Enumeration Areas used to form clusters. - Utilisant les secteurs de dénombrement.

APT = Apartment Buildings used to form clusters. - Utilisant les immeubles d'appartements.

³ ISR = Inverse Sampling Rate (after 1986 sample reduction).

FSI = Fraction de sondage inverse (après la réduction de la taille de l'échantillon en 1986)



British Columbia

Table 9.2

Non Self-Representing Areas

Economic Region	Stratum	Stratum Type	LF Population ¹	Number of PSUs	Number selected	ISR ²
Région économique	Strate	Type de strate	Population active ¹	Nombre de UPE	Nombre sélectionné	FSI ²
910	91000	urban/urbaine	9962	4	2	45
	91020	rural/rurale	11790	7	2	45
920	92000	urban/urbaine	28427	7	2	100
	92020	rural/rurale	41267	16	3	100
930	93000	mixed/mixte	60241	16	3	180
940	94000	mixed/mixte	37801	13	4	97.5
950	95000	rural/rurale	26526	10	2	180
960	96000	urban/urbaine	59360	13	3	180
	96020	rural/rurale	35901	16	2	180
	96040	rural/rurale	44352	16	2	180
970	97000	mixed/mixte	80805	18	4	135
980	98080	EA dsgn/SD	14165	1	1	140
TOTAL	12	-	450597	137	30	-

¹ Non-institutional civilian population 15 years and over. - Personnes civiles non pensionnaires d'institution âgées de 15 ans et plus.

² Inverse sampling rate. - Fraction de sondage inverse.

ER CD COMPOSITION
RE COMPOSITION EN DR

910	Division 01
920	Divisions 03,39
930	Divisions 05,07,35,37
940	Divisions 31,33
950	Divisions 09,11,13,15,27,29
960	Divisions 17,19,21,23,25,43
970	Divisions 41,51,53
980	Division 55
990	Divisions 45,47,49,57 (REMOTE and SRU ONLY / éloignées et UAR seulement)

Table 9.3

Self-Representing Area List

Tableau 9.3

Liste des unités autoreprésentatives

Serial No. ¹	Name Of SRU	1981 LF Population	No. Of Sub-Units	Numerical Identification	Economic Region	Cluster Design ²	Design ISR ³
No. de Série ¹	Nom de l'UAR	Population active de 1981	No. de sous-unités	Code d'identification	Région économique	Formation des grappes ²	FSI selon le plan ³
1	Cranbrook	11631	1	91101	910	VR	70
2	Kimberley	5616	1	91201	910	VR	70
3	Kelowna	58940	5	93101	930	BF/EA	165
4	Vernon	20191	2	93201	930	VR	165
5	Penticton	18542	1	93301	930	VR	165
6	Trail/Rossland	17825	1	93401	930	VR	165
7	Kamloops	47406	6	94101	940	BF/EA	100
8	Vancouver	1005675	53	95101	950	BF/EA	279 (419)
	Vancouver			95501	950	BF/EA	279 (419)
	" APT		2	95100	950	APT	279 (419)
	" APT			95500	950	APT	279 (419)
9	Mission/Matsqui	54597	4	95201	950	VR	165
10	Chilliwack	30225	2	95301	950	VR	120
11	Powell River	12493	1	95401	950	VR	165
12	Nanaimo	36571	3	96201	960	VR	140
13	Victoria	184511	17	96101	960	BF/EA	170 (189)
	" APT		1	96100	960	APT	170 (189)
14	Port Alberni	16448	1	96301	960	VR	165
15	Prince George	48354	5	97101	970	BF/EA	140
16	Dawson Creek/Ft. St. John	18526	2	98101	980	VR	70
17	Kitimat/Terrace	19703	1	99201	990	VR	70
18	Prince Rupert	12552	1	99101	990	VR	70
	TOTAL	1619806	107	REGULAR			
			3	APT			

¹ As on the map. - Voir la carte

² BF = Block-face counts from Area Master File used to form clusters. - Utilisant les chiffres pour les côtés d'îlots provenant du fichier principal des régions.

VR = Counts derived from Visitation Records used to form clusters. - Utilisant les chiffres provenant des registres des visites

EA = Enumeration Areas used to form clusters. - Utilisant les secteurs de dénombrement

APT = Apartment Buildings used to form clusters. - Utilisant les immeubles d'appartements

ISR = Inverse Sampling Rate (after 1986 sample reduction).

FSI = Fraction de sondage inverse (après la réduction de la taille de l'échantillon en 1986)

Sample Summary

Table 10A

Sommaire de l'échantillon

Tableau 10A

Province	Design CV	Non Self-Representing				Self-Representing		
		Non autoreprésentatif				Autoreprésentatif		
	CV selon plan	1981 LF Population	Number of Strata	Number of PSUs	Number of PSUs selected	1981 LF Population	Number of Sub-Units	Number of APT Sub-Units
		Population active en 1981	Nombre de strata	Nombre de de UPE	Nombre de de UPE sélectionnés	Population active en 1981	Nombre de sous-unités	Nombre de sous- unités APT
Nfld. – T.-N.	7.30	19134	13	107	26	197136	32	0
P.E.I. – I.-P.-É.	8.90	62381	1	12	12	27390	13	0
N.S. – N.-É.	5.44	271491	15	148	33	355123	48	1
N.B. – N.-B.	6.44	251488	16	159	32	257027	61	0
Que.	3.76	1293034	18	276	41	3584785	168	8
Ont.	3.69	1376563	23	305	47	5192518	215	21
Man.	6.75	216461	15	137	31	524134	68	2
Sask.	7.26	336447	20	225	40	352140	91	1
Alb. – Alta	6.20	504034	20	198	52	1118006	125	4
B.C. – C.-B.	5.39	450597	12	137	30	1619806	107	3
Canada		4954430	153	1704	344	13228065	928	40

Sample Summary Following 1989-1990 Increase

Table 10B

Sommaire de l'échantillon après l'augmentation en 1989-1990

Tableau 10B

Province	Household			Sampling Fraction (%)	Actual CV (%) Unemployed
	Menages				
	Self-Representing	Non-Self-Representing	Total		
	Autoreprésentatif	Non-auto-représentatif		Fraction de sondage (%)	CV actuel (%) chômeurs
Newfoundland – Terre-Neuve	1263	1707	2970	1.73	7.4
Prince Edward Island – Île-du-Prince-Édouard	411	1010	1421	3.21	6.9
Nova Scotia – Nouvelle-Écosse	1850	2559	4409	1.39	5.5
New Brunswick – Nouveau-Brunswick	1923	1920	3843	1.54	5.5
Quebec	7053	4903	11956	0.47	3.9
Ontario	12126	5919	18045	0.51	3.6
Manitoba	2080	2217	4297	1.07	7.1
Saskatchewan	2609	2354	4963	1.34	6.1
Alberta	3096	2529	5625	0.62	5.1
British Columbia – Colombie-Britannique	3641	1829	5470	0.45	5.8
Canada	36052	26947	62999	0.64	1.9

Special Area Sample

Table 11

L'échantillon dans les secteurs spéciaux

Tableau 11

Province	Frame	Coverage	Design
	Base	Couverture	Plan
Newfoundland – Terre-Neuve	instit & milit remote/éloignées	province ER/RE 30	collective EAs/SD collectifs place names/noms de localités
Nova Scotia – Nouvelle-Écosse	instit & milit	province	collective EAs/SD collectifs
New Brunswick – Nouveau-Brunswick	instit & milit	province	collective EAs/SD collectifs
Quebec	instit & milit remote/éloignées remote/éloignées	province ER/RE 400,490 ER/RE 400,480,490	collective EAs/SD collectifs place names/noms de localités urban centres/centres urbains
Ontario	instit & milit remote/éloignées	province ER/RE 591,592	collective EAs/SD collectifs EAs/SD
Manitoba	instit & milit remote/éloignées	province ER/RE 650,680	collective EAs/SD collectifs place names/noms de localités
Saskatchewan	instit & milit remote/éloignées	province ER/RE 750,760	collective EAs/SD collectifs place names/noms de localités
Alberta	instit & milit remote/éloignées	province ER/RE 870,880	collective EAs/SD collectifs EAs/SD
British Columbia – Colombie-Britannique	instit & milit remote/éloignées	province ER/RE 980,990	collective EAs/SD collectifs place names/noms de localités, EAs/SD



CLUSTER DIAGRAM SCHÉMA DE LA GRAPPE

1 01

PSL
JPE 2 8 3 1 4 2

GROUP
GROUPE 3 6 5

CLUSTER
GRAPPE 4 0 1 8

ROTATION NO.
N° DE RENOU-
VELLEMENT

5 4

COMPTÉ 6 0 4 8

SURVEY
IDENTIFICATION
IDENTIFICATION
DE L'ENQUÊTE

7 1

RANDOM START
ORIGINE CHOISIE
AU HASARD

8 0 0 3

SAMPLING RATIO
FRACTION DE
SONDAGE

9 0 1 3

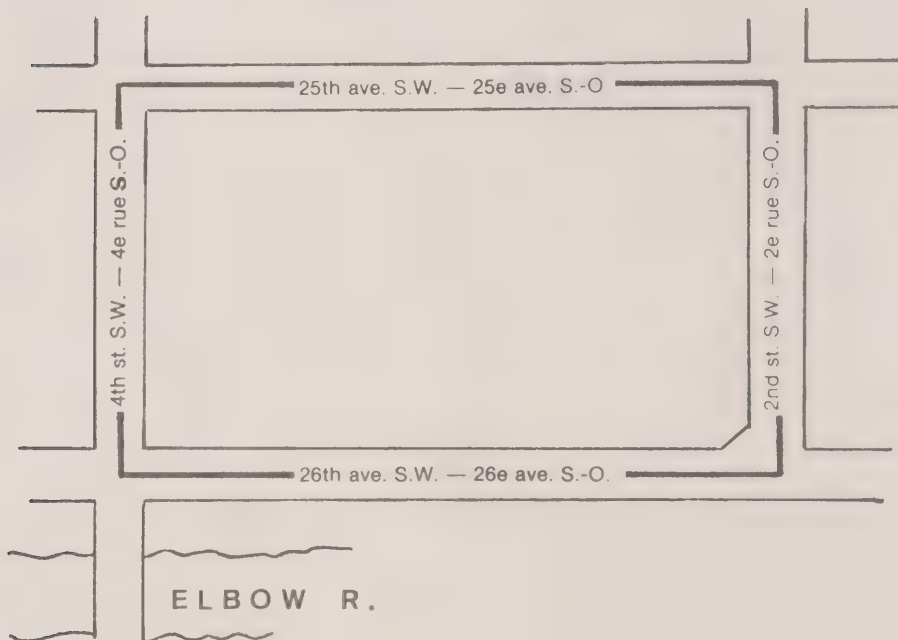
DATE OF
INTRODUCTION
DATE
D'INTRODUCTION

10 1 0 8 4
MO YR
MO AN

OVERLAP
CLUSTER

CHEVAUCHEMENT
N° DE GRAPPE

11 - - - - -



CALGARY

0 3 1 . 0 0



Statistics Canada

Statistique Canada

CLUSTER DIAGRAM SCHÉMA DE LA GRAPPE

1 01

PSU
TYPE 2 2 4 0 2 9GROUP
GROUPE 3 2 3CLUSTER
GRAPPE 4 0 0 0ROTATION NO
N° DE RENOU
VELLEMENT 5 2COUNT
COMPTÉ 6 0 8 1SURVEY
IDENTIFICATION
IDENTIFICATION
DE L'ENQUÊTE 7 1RANDOM START
ORIGINE CHOISIE
AU HASARD 8 0 0 2SAMPLING RATIO
FRACTIONNEL
SONDAGE 9 0 1 0DATE OF
INTRODUCTION
DATE
D'INTRODUCTION 10 0 2 8 5OVERLAP
CLUSTER
CHEVAUCHEMENT
N° DE GRAPPE 11LOCATION
ENDROIT

LUNENBURG MUN

RUR

Appendix F

Annexe F



Statistics
Canada

Statistique
Canada

HOUSEHOLD RECORD DOCKET

CONFIDENTIAL
when completed

1 FORM NO. **03**

Docket no. **2** Survey date **3** Assignment no. **4** Designated interviewer no. **5** Your interviewer no. **6**

P.S.U. **6** Group **7** Cluster **8** Listing **9** Mult. **10** Type of dwelling **11**

Record time of every call on this household

Listing address **12**

11 INTERVIEWER CHECK ITEM:
• Is this the first interview at this dwelling or a new household since last interview?
Yes ☐ Go to 12 No ☐ Go to 20

12 IS (read listing address in item 10) YOUR CORRECT MAILING ADDRESS?
Yes ☐ Go to 13 No ☐ Enter correct mailing address on Form 23 and go to 13

FIRST INTERVIEW WITH THIS HOUSEHOLD

13 WOULD YOU PREFER TO BE INTERVIEWED IN ENGLISH OR IN FRENCH?
English ☐ French ☐ Either ☐ Neither ☐

14 INTERVIEWER CHECK ITEM:
Language of interview
English ☐ French ☐ Other ☐ Go to 15

15 WHAT ARE THE NAMES OF ALL PERSONS NOW LIVING OR STAYING HERE WHO HAVE NO USUAL PLACE OF RESIDENCE ELSEWHERE?
Enter names in 32

16 ARE THERE ANY PERSONS AWAY FROM THIS HOUSEHOLD ATTENDING SCHOOL, VISITING, TRAVELLING OR IN HOSPITAL WHO USUALLY LIVE HERE?
Yes ☐ Enter names in 32 and go to 22 No ☐ Go to 22

22 DOES ANYONE ELSE LIVE AT THIS DWELLING SUCH AS OTHER RELATIVES, ROOMERS, BOARDERS OR EMPLOYEES?
Yes ☐ Enter names in 32, COMPLETE 33 through 40 and go to 42 No ☐ • For a first interview, COMPLETE 33 through 40 and go to 42 • For a subsequent interview, go to 42

20 SUBSEQUENT INTERVIEW

• If Personal interview ☐ Go to 21
• Otherwise ask ☐ Go to 21

21 ARE YOU STILL LIVING IN THE SAME DWELLING AS LAST MONTH?
• where possible, state listing address
Yes ☐ Go to 21 No ☐ Personal visit required

21 DO THE FOLLOWING PERSONS STILL LIVE OR STAY AT THIS DWELLING?
• Read all names in 32
• Enter appropriate code in 40

HRD Pg Ln	Names of household members	Age	Sex	M S	F a m i l y	R to H	Educ	Membership	Answers to supplementary questions
									A B C D E F G H I J K L M
1	Given name								
1	Surname								
2	Given name								
2	Surname								
3	Given name								
3	Surname								
4	Given name								
4	Surname								
5	Given name								
5	Surname								
6	Given name								
6	Surname								
7	Given name								
7	Surname								
8	Given name								
8	Surname								

42 IS THIS DWELLING OWNED BY A MEMBER OF THIS HOUSEHOLD?
Yes ☐ Go directly to FORMS 05 & 06 No ☐ Complete FORM 04

COMPLETE AT END OF INTERVIEW

43 FOR ALL HOUSEHOLDS

Telephone no. **43** Permission to interview by phone: Granted ☐ Denied ☐

No telephone **43** Telephone no. 2 refused **43** Go to 44

44 Determine and record the best time to call on this household.

45 RESPONSE CODE
Month **45** Code **45**

46 Was this interview conducted by telephone?
Yes ☐ No ☐

47 Forms Control
Printed **47** Completed **47**

48 NOTES
See over for additional NOTES

Authority - Statistics Act, Revised Statutes of Canada 1985, chapter S-19



LABOUR FORCE SURVEY QUESTIONNAIRE

CONFIDENTIAL when completed

Docket No 2 _____ Survey date 3 _____ Assignment No 4 _____

HFD page - line No 5 _____ Given name 6 _____ Surname 7 _____ 1 FORM NO 05

10 LAST WEEK, DID ... WORK AT A JOB OR BUSINESS? (Regardless of the number of hours.)
Yes ☐ No ☐ Go to 30
PERMANENTLY unable to work ☐ Go to 50

11 DID ... HAVE MORE THAN ONE JOB OR BUSINESS LAST WEEK?
Yes ☐ No ☐ Go to 13

12 WAS THIS A RESULT OF CHANGING EMPLOYERS LAST WEEK?
Yes ☐ No ☐

13 HOW MANY HOURS PER WEEK DOES ... USUALLY WORK AT HIS/HER:
(Main) JOB? If total 30 or more go to 15
Other jobs?

14 WHAT IS THE REASON ... USUALLY WORKS LESS THAN 30 HOURS PER WEEK?
 Enter code

15 LAST WEEK, HOW MANY HOURS OF OVERTIME OR EXTRA HOURS DID ... WORK?
(Include paid and unpaid time at all jobs) If none enter 00

16 LAST WEEK, HOW MANY HOURS WAS AWAY FROM WORK FOR ANY REASON? (Holiday, vacation, illness, labour dispute, etc.)
(From all jobs) If none enter 00 and go to 18

17 WHAT WAS THE MAIN REASON FOR BEING AWAY FROM WORK?
 Enter code

18 HOW MANY HOURS DID ... ACTUALLY WORK LAST WEEK AT HIS/HER:
(Main) JOB?
Other jobs?

19 IN THE PAST 4 WEEKS, HAS ... LOOKED FOR ANOTHER JOB?
Yes ☐ No ☐ Go to 72

20 LAST WEEK, DID ... HAVE A JOB OR BUSINESS AT WHICH HE/SHE DID NOT WORK?
Yes ☐ Go to 33 No ☐ Go to 50

21 LAST WEEK, DID ... HAVE A JOB TO START AT A DEFINITE DATE IN THE FUTURE?
Yes ☐ No ☐ Go to 50

22 COUNTING FROM THE END OF LAST WEEK, IN HOW MANY WEEKS WILL ... START TO WORK AT HIS/HER NEW JOB?
 Go to 50

23 WHY WAS ... ABSENT FROM WORK LAST WEEK?
 Enter code and if code 6 go to 32

24 DID ... HAVE MORE THAN ONE JOB OR BUSINESS LAST WEEK?
Yes ☐ No ☐

25 HOW MANY HOURS PER WEEK DOES ... USUALLY WORK AT HIS/HER:
(Main) JOB? If total 30 or more go to 37
Other jobs?

26 WHAT IS THE REASON ... USUALLY WORKS LESS THAN 30 HOURS PER WEEK?
 Enter code

27 UP TO THE END OF LAST WEEK, HOW MANY WEEKS HAS ... BEEN CONTINUOUSLY ABSENT FROM WORK?

28 IS ... GETTING ANY WAGES OR SALARY FROM HIS/HER EMPLOYER FOR ANY TIME OFF LAST WEEK?
Yes ☐ No ☐

29 INTERVIEWER CHECK ITEM:
* If code 5 (layoff) in 33 ☐ go to 56
* Otherwise ☐ go to 40

30 LAST WEEK, DID ... HAVE A JOB OR BUSINESS AT WHICH HE/SHE DID NOT WORK?
Yes ☐ No ☐ Go to 55

31 WHEN DID ... LAST WORK AT A JOB OR BUSINESS?
No change ☐ or ☐ Mo Yr If month unknown, enter - in month

32 INTERVIEWER CHECK ITEM:
(1) # 51 is before go to 55
(2) # 51 is equal to or later than go to 53

33 DID ... USUALLY WORK 30 OR MORE HOURS PER WEEK?
Full-time ☐ Part-time ☐ (Less than 30 hours per week)

34 WHAT WAS THE MAIN REASON WHY ... LEFT THAT JOB?
 Enter code

35 INTERVIEWER CHECK ITEM:
* If "perm. unable to work" in 10 ☐ go to 80
* Otherwise ☐ go to 56

36 IN THE PAST 6 MONTHS, HAS ... LOOKED FOR WORK?
Yes ☐ No ☐ Go to 64

37 * IN THE PAST 4 WEEKS, WHAT HAS ... DONE TO FIND WORK? Mark all methods reported
Nothing ☐ Go to 62
* IN THE PAST 4 WEEKS, HAS ... DONE ANYTHING ELSE TO FIND WORK? Mark all other methods reported. For each method given, ask:
* WHEN DID ... LAST ... (Repeat method) ?
Method used No. of weeks ago (and any week)

38 Checked with:
PUBLIC employment AGENCY ☐
PRIVATE employment AGENCY ☐
UNION ☐
EMPLOYERS directly ☐
FRIENDS or relatives ☐
Placed or answered ADS ☐
LOOKED at job ADS ☐
OTHER, Specify in NOTES ☐

39 UP TO THE END OF LAST WEEK, HOW MANY WEEKS HAS ... BEEN LOOKING FOR WORK? (Not counting weeks worked.)

40 WHAT WAS ...'S MAIN ACTIVITY BEFORE HE/SHE STARTED TO LOOK FOR WORK?
 Enter code

41 IS ... LOOKING FOR A JOB TO LAST MORE THAN 6 MONTHS?
Yes ☐ No ☐ (More than 6 months) (6 months or less)

42 IS ... LOOKING FOR A FULL-TIME OR PART-TIME JOB?
Full-time ☐ Part-time ☐ (30 or more hours per week) (Less than 30 hours per week)
Go to 63

43 WHAT WAS THE MAIN REASON WHY ... DID NOT LOOK FOR WORK LAST WEEK?
 Enter code

44 WAS THERE ANY REASON WHY ... COULD NOT TAKE A JOB LAST WEEK?
 Enter code

45 INTERVIEWER CHECK ITEM:
* If "No" (never worked) in 50 ☐ go to 80
* If upper circle in 52 is marked ☐ go to 80
* Otherwise ☐ go to 72

46 EDUCATIONAL ACTIVITIES: If age 65 or over, go to 80.
LAST WEEK, WAS ... ATTENDING A SCHOOL, COLLEGE OR UNIVERSITY?
Yes ☐ No ☐ Go to 80

47 WAS ... ENROLLED AS A FULL-TIME OR A PART-TIME STUDENT?
Full time ☐ Part time ☐

48 WHAT KIND OF SCHOOL WAS THIS?
 Enter code

49 INFORMATION SOURCE:
HFD page-line No. of person providing the above information:
Last interview This interview
Authority: Statistics Act Chapter 15, Statutes of Canada 1970-71-72.

72 FOR WHOM DID ... WORK? (Name of business, government dept. or agency, or person)
No change ☐ or ☐

73 WHEN DID ... START WORKING FOR THIS EMPLOYER?
No change ☐ or ☐ Mo Yr If month unknown, enter - in month

74 WHAT KIND OF BUSINESS, INDUSTRY OR SERVICE WAS THIS? (Give full description: e.g., federal government, canning industry, forestry services.)
No change ☐ or ☐

75A WHAT KIND OF WORK WAS ... DOING? (Give full description: e.g., office clerk, factory worker, forestry technician.)
No change ☐ or ☐

75B IN THIS WORK, WHAT WERE ... MOST IMPORTANT ACTIVITIES OR DUTIES? (Give full description: e.g., filing documents, drying vegetables, forest examiner.)
No change ☐ or ☐

76 Class of worker:
Main job ☐ Enter code
Other job ☐ Enter code go to 80

77 NOTES



DOSSIER DU MÉNAGE

DOCUMENT CONFIDENTIEL
une fois rempli

1 FORMULE

03

N° du dossier 2 Date d'enquête 3 N° de tâche 4 N° de l'intervieweur désigné 5 Votre n° d'intervieweur 6 U.P.E. 7 Groupe 8 Genre de logement 9 Notes l'heure de chaque visite ou appel à ce ménage 10 Adresse de liste 11 À L'INTERVIEWEUR: 12 EST-CE QUE (lisez l'adresse de liste au poste 10) EST VOTRE ADRESSE POSTALE EXACTE? 13 PRÉFÉREZ-VOUS ÊTRE INTERVIEWÉ(E) EN FRANÇAIS OU EN ANGLAIS? 14 À L'INTERVIEWEUR: 20 INTERVIEW SUBSÉQUENTE 21 LES PERSONNES SUIVANTES DEMEURENT-ELLES ENCORE DANS CE LOGEMENT? 22 Y A-T-IL QUELQU'UN D'AUTRE QUI DEMEURE DANS CE LOGEMENT, PAR EXEMPLE UN PARENT, UN CHAMBREUR, UN PENSIONNAIRE OU UN EMPLOYÉ?

31	32	33	34	35	36	37	38	40	50	Réponses aux questions supplémentaires											
DM	Noms des membres du ménage		Age	S	E	F	N	Composition													
P				e	x	m	e		A	B	C	D	E	F	G	H	I	J	K	L	M
1	Prénom	Nom de famille							1												
2	Prénom	Nom de famille							2												
3	Prénom	Nom de famille							3												
4	Prénom	Nom de famille							4												
5	Prénom	Nom de famille							5												
6	Prénom	Nom de famille							6												
7	Prénom	Nom de famille							7												
8	Prénom	Nom de famille							8												

42 CE LOGEMENT APPARTIENT-IL À UN MEMBRE DE CE MÉNAGE? Oui 1/ Passez directement aux FORMULES 05 et 06 Non 2/ Remplissez la FORMULE 04

43 POUR TOUS LES MÉNAGES N° de téléphone 1/ ou 2/ N° de téléphone 2/ refusé 3/ Accordée 4/ Refusée

44 Déterminez et inscrivez le meilleur moment pour contacter ce ménage

45 CODE-RÉPONSE Mois Code

46 Cette interview a-t-elle été faite par téléphone? Oui 1/ Non 2/

47 Contrôle des formules Formules 04 05 06 Imprimées Remplies

48 NOTES



Statistique Canada

QUESTIONNAIRE DE L'ENQUÊTE SUR LA POPULATION ACTIVE

DOCUMENT CONFIDENTIEL - une fois rempli

N° de classeur 2 _____ Date d'enquête 3 _____ N° de tâche 4 _____
N° de page-ligne du DM Prénom _____ Mois _____ Année _____ Nom de famille _____ 1 FORMULE 05
5 _____ 6 _____ 7 _____

10 LA SEMAINE DERNIÈRE, ... A-T-IL(ELLE) TRAVAILLÉ À UN EMPLOI OU À UNE ENTREPRISE? (Peu importe le nombre d'heures.)
Oui ☐ Non ☐ Passer à 30
Inscrite PERMANENTE ☐ Passer à 50

11 AVAIT-IL(ELLE) PLUS D'UN EMPLOI OU PLUS D'UNE ENTREPRISE LA SEMAINE DERNIÈRE?
Oui ☐ Non ☐ Passer à 13

12 ÉTAIT-CE DÙ À UN CHANGEMENT D'EMPLOYEUR LA SEMAINE DERNIÈRE?
Oui ☐ Non ☐

13 COMBIEN D'HEURES PAR SEMAINE TRAVAILLÉ-T-IL(ELLE) HABITUELLEMENT:
À SON EMPLOI (principal)? ☐ Si le total est 30 ou plus, passer à 15
à d'autres emplois? ☐

14 POURQUOI ... TRAVAILLÉ-T-IL(ELLE) HABITUELLEMENT MOINS DE 30 HEURES PAR SEMAINE?
☐ Inscrivez le code

15 LA SEMAINE DERNIÈRE, COMBIEN D'HEURES SUPPLÉMENTAIRES ... A-T-IL(ELLE) TRAVAILLÉ?
(Compter les heures supplémentaires ou non, à tous les emplois) ☐ Pour "aucune", inscrire 00

16 LA SEMAINE DERNIÈRE, COMBIEN D'HEURES ... A-T-IL(ELLE) ÉTÉ ABSENT(E) DU TRAVAIL POUR UNE RAISON QUELCONQUE? (Jour férié, vacances, maladie, conflit de travail, etc.)
(Tous les emplois) ☐ Inscrivez 00 et passer à 18

17 QUELLE ÉTAIT LA RAISON PRINCIPALE DE CETTE ABSENCE?
☐ Inscrivez le code

18 COMBIEN D'HEURES ... A-T-IL(ELLE) EFFECTIVEMENT TRAVAILLÉ LA SEMAINE DERNIÈRE:
À SON EMPLOI (principal)? ☐
à d'autres emplois? ☐

19 AU COURS DES 4 DERNIÈRES SEMAINES, S'EST-IL(ELLE) CHERCHÉ UN AUTRE EMPLOI?
Oui ☐ Non ☐ Passer à 72

30 LA SEMAINE DERNIÈRE, ... AVAIT-IL(ELLE) UN EMPLOI OU UNE ENTREPRISE AUXQUELS IL(ELLE) N'A PAS TRAVAILLÉ?
Oui ☐ Passer à 33 Non ☐ Passer à 50

31 LA SEMAINE DERNIÈRE, ... AVAIT-IL(ELLE) UN EMPLOI DEVANT COMMENCER À UNE DATE FUTURE DÉTERMINÉE? Oui ☐ Non ☐ Passer à 50

32 À PARTIR DE LA FIN DE SEMAINE DERNIÈRE, DANS COMBIEN DE SEMAINES ... DOIT-IL(ELLE) COMMENCER À TRAVAILLER À SON NOUVEL EMPLOI?
☐ Passer à 50

33 POURQUOI ... S'EST-IL(ELLE) ABSENT(E) DU TRAVAIL LA SEMAINE DERNIÈRE?
☐ Inscrivez le code. Si c'est 6, passer à 32

34 ... AVAIT-IL(ELLE) PLUS D'UN EMPLOI OU PLUS D'UNE ENTREPRISE LA SEMAINE DERNIÈRE?
Oui ☐ Non ☐

35 COMBIEN D'HEURES PAR SEMAINE TRAVAILLÉ-T-IL(ELLE) HABITUELLEMENT:
À SON EMPLOI (principal)? ☐ Si le total est 30 ou plus, passer à 37
à d'autres emplois? ☐

36 POURQUOI ... TRAVAILLÉ-T-IL(ELLE) HABITUELLEMENT MOINS DE 30 HEURES PAR SEMAINE?
☐ Inscrivez le code

37 JUSQU'À LA FIN DE LA SEMAINE DERNIÈRE, PENDANT COMBIEN DE SEMAINES CONSECUTIVES ... AVAIT-IL(ELLE) ÉTÉ ABSENT(E) DU TRAVAIL?
☐

38 ... REÇOIT-IL(ELLE) UNE RÉMUNÉRATION DE SON EMPLOYEUR POUR SES ABSENSES DE LA SEMAINE DERNIÈRE?
Oui ☐ Non ☐

39 À L'INTERVIEWER:
* Si code 5 (mise à pied) à 33 ☐ passer à 56
* Autrement ☐ passer à 40

40 AU COURS DES 4 DERNIÈRES SEMAINES, S'EST-IL(ELLE) CHERCHÉ UN AUTRE EMPLOI?
Oui ☐ Non ☐ Passer à 72

DESCRIPTION DE L'EMPLOI OU DE L'ENTREPRISE (activité principale)

72 POUR QUI ... TRAVAILLÉ-T-IL(ELLE)? (Nom de l'entreprise, de l'organisme gouvernemental ou de la personne)
Autre changement ☐ ou ☐

73 QUAND ... A-T-IL(ELLE) COMMENCÉ À TRAVAILLER POUR CET EMPLOYEUR?

Autre changement ☐ ou ☐ Si le mois est inconnu, mettez -- à "Mois"

74 DE QUEL GENRE D'ENTREPRISE, D'INDUSTRIE OU DE SERVICE S'AGISSAIT-IL? (Donnez une description complète, par ex. gouvernement fédéral, conserverie, services forestiers.)

Autre changement ☐ ou ☐

75A QUEL GENRE DE TRAVAIL ... FAISAIT-IL(ELLE)? (Par exemple, commis de bureau, travailleur(se) d'usine, technicien(ne) forestier(ère).)

Autre changement ☐ ou ☐

75B DANS CE TRAVAIL, QUELLES ÉTAIENT LES ACTIVITÉS OU FONCTIONS LES PLUS IMPORTANTES DE ...? (Par ex., classement de documents, séchage de légumes, estimation forestière.)

Autre changement ☐ ou ☐

76 Catégorie de travailleur:
Emploi principal ☐ Aut. changement ☐ ou ☐ Inscrivez le code

77 Autre emploi ☐ Aut. changement ☐ ou ☐ Inscrivez le code passer à 80

NOTES
N° de page-ligne du DM de la personne qui a fourni les renseignements ci-dessus.
1-5030-4/5 4-6-86

50 ... A-T-IL(ELLE) DÉJÀ TRAVAILLÉ À UN EMPLOI OU À UNE ENTREPRISE? Oui ☐ Non ☐ Passer à 55

51 QUAND ... A-T-IL(ELLE) TRAVAILLÉ À UN EMPLOI OU À UNE ENTREPRISE LA DERNIÈRE FOIS?
Aut. changement ☐ ou ☐ Si le mois est inconnu, mettez -- à "Mois"

52 À L'INTERVIEWER:
(1) Si le mois à 51 se situe avant ☐ 1 ☐ Passer à 55
(2) Si le mois à 51 est le même que celui se situant après ☐ 2 ☐ Passer à 53

53 TRAVAILLÉ-T-IL(ELLE) HABITUELLEMENT 30 HEURES OU PLUS PAR SEMAINE?
À temps plein ☐ À temps partiel ☐ (moins de 30 heures par semaine)

54 QUELLE ÉTAIT LA RAISON PRINCIPALE POUR LAQUELLE ... A LAISSÉ CET EMPLOI?
☐ Inscrivez le code

55 À L'INTERVIEWER:
* Si "Incapable permanentement" à 10 ☐ passer à 80
* Autrement ☐ passer à 56

56 AU COURS DES 6 DERNIERS MOIS, ... S'EST-IL(ELLE) CHERCHÉ DU TRAVAIL?
Oui ☐ Non ☐ Passer à 64

57 * AU COURS DES 4 DERNIÈRES SEMAINES, QU'A FAIT ... POUR SE TROUVER DU TRAVAIL? Notez toute méthode indiquée.
Rien ☐ Passer à 62

* AU COURS DES 4 DERNIÈRES SEMAINES, ... A-T-IL(ELLE) FAIT AUTRE CHOSE POUR SE TROUVER DU TRAVAIL? Notez toute autre méthode indiquée. Pour chaque méthode donnée, demandez:
* QUAND ... LA DERNIÈRE FOIS? (Répétez la méthode) Nombre de semaines depuis la dernière fois (exacte la semaine d'int.)

S'est-il(elle) adressé(e) à un BUREAU de placement PUBLIC ☐ 2
à un BUREAU de placement PRIVÉ ☐ 3
à un SYNDICAT ☐ 4
directement à des EMPLOYEURS ☐ 5
à des AMIS ou des parents ☐ 6
A-t-il(elle) publié une ANNONCE ou répondu à une ANNONCE ☐ 7
A-t-il(elle) CONSULTÉ les ANNONCES ☐ 8
AUTRE. Précisez dans les NOTES ☐ 9

58 JUSQU'À LA FIN DE LA SEMAINE DERNIÈRE, PENDANT COMBIEN DE SEMAINES ... AVAIT-IL(ELLE) CHERCHÉ DU TRAVAIL? (Ne comptant pas les semaines travaillées.)

59 QUELLE ÉTAIT L'ACTIVITÉ PRINCIPALE DE ... AVANT QU'IL(ELLE) COMMENCÉ À SE CHERCHER DU TRAVAIL?
☐ Inscrivez le code

60 CHERCHE-T-IL(ELLE) UN EMPLOI DE PLUS DE 6 MOIS?
Oui ☐ Non ☐ (Plus de 6 mois) (Moins de 6 mois incl. 6 mois)

61 ... CHERCHE-T-IL(ELLE) UN EMPLOI À TEMPS PLEIN OU UN EMPLOI À TEMPS PARTIEL?
À temps plein ☐ À temps partiel ☐ (30 heures ou plus par semaine) (moins de 30 heures par semaine)

62 QUELLE ÉTAIT LA RAISON PRINCIPALE POUR LAQUELLE ... N'A PAS CHERCHÉ DE TRAVAIL LA SEMAINE DERNIÈRE?
☐ Inscrivez le code

63 Y A-T-IL UNE RAISON QUI AURAIT EMPÊCHÉ ... DE PRENDRE UN EMPLOI LA SEMAINE DERNIÈRE?
☐ Inscrivez le code

64 À L'INTERVIEWER:
* Si "Non" (jamais travaillé) à 50 ☐ passer à 80
* S'il y a un "X" dans le cercle du haut à 52 ☐ passer à 80
* Autrement ☐ passer à 72

ACTIVITÉ SCOLAIRE (15-65 ans et plus, passant à 90)

80 LA SEMAINE DERNIÈRE, ... FRÉQUENTAIT-IL(ELLE) UNE ÉCOLE, UN COLLÈGE OU UNE UNIVERSITÉ? Oui ☐ Non ☐ Passer à 90

81 ... ÉTAIT-IL(ELLE) INSCRIT(E) COMME ÉTUDIANT(E) À TEMPS PLEIN OU À TEMPS PARTIEL?
À temps plein ☐ À temps partiel ☐

82 DE QUEL GENRE D'ÉCOLE S'AGIT-IL?
☐ Inscrivez le code

SOURCE DE RENSEIGNEMENTS

90 N° de page-ligne du DM de la personne qui a fourni les renseignements ci-dessus.
La dernière interview ☐ Cette interview ☐
Déclaration exigée en vertu de la Loi sur la statistique, chapitre 15, Statuts du Canada de 1970-71-72.

SEE REVERSE FOR ENGLISH

Appendix G

Definitions and Explanations

Labour Force

The labour force is composed of those members of the civilian non-institutional population 15 years of age and over who, during the reference week, were employed or unemployed.

Employed

Employed persons are those who, during the reference week:

- (a) did any work¹ at all
- (b) had a job but were not at work due to:
 - own illness or disability
 - personal or family responsibilities
 - bad weather
 - labour dispute
 - vacation
 - other reason not specified above (excluding persons on layoff and persons whose job attachment was a job to start at a definite date in the future).

Unemployed

Unemployed persons are those who, during the reference week:

- (a) were without work, had actively looked for work in the past four weeks (ending with reference week), and were available for work ²;
- (b) had not actively looked for work in the past four weeks but had been on layoff³ and were available for work;

¹ Work includes any work for pay or profit, that is, paid work in the context of an employer-employee relationship, or self-employment. It also includes unpaid family work where unpaid family work is defined as unpaid work which contributed directly to the operation of a farm, business or professional practice owned or operated by a related member of the household.

² Persons in this group meeting the following criteria are regarded as available:

- (i) were full-time students seeking part-time work who also met condition (ii) below. (Full-time students looking for full-time work are classified as not available for work in the reference week.)

- (ii) reported that there was no reason why they could not take a job in reference week, or if they could not take a job it was because of "own illness or disability", "personal or family responsibilities", or "already had a job".

³ Persons are classified as being on layoff only when they expect to return to the job from which they were laid off.

Annexe G

Définitions et explications

Population active

La population active comprend les membres de la population civile hors institution âgés de 15 ans et plus qui avaient un emploi (personnes occupées) ou étaient en chômage pendant la semaine de référence.

Personnes occupées

Les personnes occupées sont celles qui, au cours de la semaine de référence:

- a) ont fait un travail¹ quelconque
- b) avaient un emploi mais n'étaient pas au travail pour l'une des raisons suivantes:
 - maladie ou invalidité (de l'enquête)
 - obligations personnelles ou familiales
 - mauvais temps
 - conflit de travail
 - vacances
 - autre raison non précisée ci-dessus (à l'exception des personnes mises à pied et de celles qui devaient commencer à travailler à une date ultérieure déterminée).

Chômeurs

Les chômeurs sont les personnes qui, au cours de la semaine de référence:

- a) étaient sans emploi, avaient activement cherché du travail au cours des quatre dernières semaines (y compris la semaine de référence) et étaient prêtes à travailler ²;
- b) n'avaient pas activement cherché de travail au cours des quatre dernières semaines, mais avaient été mises à pied³ et étaient prêtes à travailler;

¹ On entend par "travail" tout travail fait contre rémunération ou en vue d'un bénéfice, c'est-à-dire le travail fait contre rémunération pour un employeur ou à son propre compte. Il peut également s'agir d'un travail familial non rémunéré, c'est-à-dire d'un travail non rémunéré qui contribue directement à l'exploitation d'une ferme, d'une entreprise ou d'un bureau professionnel, qui est possédé ou exploité par un membre apparenté du ménage.

² On considère comme prêtes à travailler les personnes de ce groupe qui:

- (i) étudiaient à plein temps tout en cherchant du travail à temps partiel et qui répondaient au critère (ii) ci-dessous (Les étudiants à plein temps qui cherchaient du travail à plein temps ne sont pas considérés comme ayant été prêts à travailler pendant la semaine de référence.)

- (ii) ont déclaré qu'il n'y avait aucune raison pour laquelle elles ne pouvaient prendre un emploi pendant la semaine de référence ou ne pouvaient en prendre un en raison "de maladie ou invalidité", "d'obligations personnelles ou familiales", ou "parce qu'elles avaient déjà un emploi"

³ On considère comme mises à pied les personnes qui s'attendent à réintégrer l'emploi qu'elles occupaient avant d'être mises à pied.

(c) had not actively looked for work in the past four weeks but had a new job to start in four weeks or less from the reference week, and were available for work.

Not in the Labour Force

Those persons in the civilian non-institutional population 15 years of age and over who, during the reference week, were neither employed nor unemployed.

Unemployment Rate

The unemployment rate represents the number of unemployed persons expressed as a percentage of the labour force. The unemployment rate for a particular group (age, sex, marital status, etc.) is the number unemployed in that group expressed as a percentage of the labour force for that group.

Participation Rate

The participation rate represents the labour force expressed as a percentage of the population 15 years of age and over. The participation rate for a particular group (age, sex, marital status, etc.) is the labour force in that group expressed as a percentage of the population for that group.

Employment/Population Ratio

The employment/population ratio represents the number of persons employed expressed as a percentage of the population 15 years of age and over. The employment/population ratio for a particular group (age, sex, marital status, etc.) is the number employed in that group expressed as a percentage of the population for that group.

c) n'avaient pas activement cherché de travail au cours des quatre dernières semaines, mais devaient commencer un nouvel emploi dans quatre semaines ou moins à compter de la semaine de référence, et étaient prêtes à travailler.

Inactifs

Il s'agit de la partie de la population civile hors institution âgée de 15 ans et plus qui était ni occupée ni en chômage pendant la semaine de référence.

Taux de chômage

Le taux de chômage représente le nombre de chômeurs exprimé en pourcentage de la population active. Le taux de chômage d'un groupe particulier (âge, sexe, état matrimonial, etc.) est le nombre de chômeurs dans ce groupe, exprimé en pourcentage de la population active dans le même groupe.

Taux d'activité

Le taux d'activité représente la population active exprimée en pourcentage de la population de 15 ans et plus. Le taux d'activité d'un groupe particulier (âge, sexe, état matrimonial, etc.) est la population active dans ce groupe exprimée en pourcentage de la population dans le même groupe.

Rapport emploi-population

Le rapport emploi-population représente le nombre de personnes occupées exprimé en pourcentage de la population de 15 ans et plus. Le rapport emploi-population d'un groupe particulier (âge, sexe, état matrimonial, etc.) est le nombre de personnes occupées dans ce groupe, exprimé en pourcentage de la population dans le même groupe.

References

Abowd, J.M. and Zellner, A. (1985). Estimating Gross Labor Flows. *Journal of Business and Economic Statistics*, 3, 254-283.

Bailar, B.A. (1975). The Effects of Rotation Group Bias on Estimates from Panel Surveys. *Journal of the American Statistical Association*, 70, 23-29.

Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis, Forecasting and Control*, San Francisco: Holden Day.

Choudhry, G.H. and Hidiroglou, M.A. (1988). Small Area Estimation: Some Investigations at Statistics Canada. *Bulletin of the International Statistical Institute, Proceedings of the 46th Session*, 451-468.

Choudhry, G.H. and Lee, H. (1987). Variance Estimation for the Canadian Labour Force Survey. *Survey Methodology*, 13, 147-161.

Choudhry, G.H., Lee, H. and Drew, J.D. (1985). Cost-Variance Optimization for the Canadian Labour Force Survey. *Survey Methodology*, 11, 33-50.

Choudhry, G.H. and Rao, J.N.K. (1989). Small Area Estimation Using Models that Combine Time Series and Cross-Sectional Data. *Proceedings of the Statistics Canada Symposium on Analysis of Data in Time*.

Chua, T.C. and Fuller, W.A. (1987). A Model for Multinomial Response Error Applied to Labor Flows. *Journal of the American Statistical Association*, 82, 46-51.

Dagum, E. B. (1980). *The X-11-ARIMA Seasonal Adjustment Method*. Ottawa, Statistics Canada, Catalogue No. 12-546.

Drew, J.D., Bélanger, Y. and Foy, P. (1985). Stratification in the Canadian Labour Force Survey. *Survey Methodology*, 11, 95-110.

Drew, J.D., Choudhry, G.H. and Gray, G.B. (1978). Some Methods for Updating Sample Survey Frames and their Effects on Estimation. *Survey Methodology*, 4, 225-263.

Drew, J.D., Dick, P. and Switzer, K. (1989). Development and testing of telephone survey methods for household surveys at Statistics Canada. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Drew, J.D., Royce, D.A. and van Baaren, A. (1989). Address Register Research at Statistics Canada. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Bibliographie

Abowd, J.M. et Zellner, A. (1985). Estimating Gross Labor Flows. *Journal of Business and Economic Statistics*, 3, 254-283.

Bailar, B.A. (1975). The Effects of Rotation Group Bias on Estimates from Panel Surveys. *Journal of the American Statistical Association*, 70, 23-29.

Box, G.E.P. et Jenkins, G.M. (1970). *Time Series Analysis, Forecasting and Control*, San Francisco: Holden Day.

Choudhry, G.H. et Hidiroglou, M.A. (1988). Small Area Estimation: Some Investigations at Statistics Canada. *Bulletin of the International Statistical Institute, Proceedings of the 46th Session*, 451-468.

Choudhry, G.H. et Lee, H. (1987). Estimation de la variance pour l'enquête sur la population active. *Techniques d'enquête*, 13, 157-172.

Choudhry, G.H., Lee, H. et Drew, J.D. (1985). Optimisation du coût et de la variance dans le cadre de l'enquête sur la population active du Canada. *Techniques d'enquête*, 11, 37-56.

Choudhry, G.H. et Rao, J.N.K. (1989). Estimation des données régionales à l'aide de modèles qui combinent des séries chronologiques et des données transversales. *Recueil du symposium à Statistique Canada sur l'analyse des données dans le temps*.

Chua, T.C. et Fuller, W.A. (1987). A Model for Multinomial Response Error Applied to Labor Flows. *Journal of the American Statistical Association*, 82, 46-51.

Dagum, E. B. (1980). *La méthode de désaisonnalisation X-11 ARMMI*, Ottawa, Statistique Canada, N° 12-546F au catalogue.

Drew, J.D., Bélanger, Y. et Foy, P. (1985). La stratification dans l'enquête sur la population active du Canada. *Techniques d'enquête*, 11, 109-124.

Drew, J.D., Choudhry, G.H. et Gray, G.B. (1978). Some Methods for Updating Sample Survey Frames and their Effects on Estimation. *Techniques d'enquête*, 4, 225-263.

Drew, J.D., Dick, P. et Switzer, K. (1989). Development and testing of telephone survey methods for household surveys at Statistics Canada. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Drew, J.D., Royce, D.A. et van Baaren, A. (1989). Address Register Research at Statistics Canada. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

- Drew, J.D., Singh, M.P. and Choudhry, G.H. (1982). Evaluation of Small Area Estimation Techniques for the Canadian Labour Force Survey. *Survey Methodology*, 8, 17-47.
- Fellegi, I.P., Gray, G.B. and Platek, R. (1967). The New Design of the Canadian Labour Force Survey. *Journal of the American Statistical Association*, 62, 421-453.
- Friedman, H.P. and Rubin, J. (1967). On Some Invariant Criteria for Grouping Data. *Journal of the American Statistical Association*, 62, 1159-1178.
- Ghangurde, P.D. (1982). Rotation Group Bias in the LF Estimates. *Survey Methodology*, 8, 86-101.
- Gentleman, J.F. (1988). Assessing Bias Due to Classification Error in Labour Force Survey Data. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Gurney, M. and Daly, J.F. (1965). A Multivariate Approach to Estimation in Periodic Sample Surveys. *Proceedings of the Social Statistics Section*, American Statistical Association, 242-257.
- Hartley, H.O. and Rao, J.N.K. (1962). Sampling with Unequal Probabilities and without Replacement. *Annals of Mathematical Statistics*, 33, 350-374.
- Hellerman, E. (1982). Overview of the Hellerman I & O Coding System. Draft memo, U.S. Bureau of the Census.
- Higginson, J. (1977). *Users Manual for the Decomposition Model Test*. Seasonal Adjustment and Time Series Staff, Statistics Canada, Ref. No. 77-01-001.
- Keyfitz, N. (1957). Estimates of Sampling Variance where Two Units Are Selected from Each Stratum. *Journal of the American Statistical Association*, 52, 503-510.
- Kumar, S. and Lee, H. (1983). Evaluation of Composite Estimation for the Canadian Labour Force Survey. *Survey Methodology*, 9, 178-201.
- Lemaître, G.E. (1988). The Measurement and Analysis of Gross Flows. Working Paper No. SSMD-88-1E. Statistics Canada.
- Lemaître, G.E. and Dufour, J. (1987). An Integrated Method for Weighting Persons and Families. *Survey Methodology*, 13, 199-207.
- Lothian, J. and Morry, M. (1977). The Problem of Aggregation Direct or Indirect Seasonal Adjustment. Research Paper, Seasonal Adjustment and Time Series Staff, Statistics Canada.
- Morry, M. (1975). A Test for Model Selection. Seasonal Adjustment and Time Series Staff, Statistics Canada, Research Paper No. 75-12-016.
- Drew, J.D., Singh, M.P. et Choudhry, G.H. (1982). Évaluation des techniques d'estimation pour les petites régions dans l'enquête sur la population active du Canada. *Techniques d'enquête*, 8, 19-52.
- Fellegi, I.P., Gray, G.B. et Platek, R. (1967). The New Design of the Canadian Labour Force Survey. *Journal of the American Statistical Association*, 62, 421-453.
- Friedman, H.P. et Rubin, J. (1967). On Some Invariant Criteria for Grouping Data. *Journal of the American Statistical Association*, 62, 1159-1178.
- Ghangurde, P.D. (1982). Le biais de renouvellement de l'échantillon dans les estimations de l'EPA. *Techniques d'enquête*, 8, 94-111.
- Gentleman, J.F. (1988). Assessing Bias Due to Classification Error in Labour Force Survey Data. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Gurney, M. et Daly, J.F. (1965). A Multivariate Approach to Estimation in Periodic Sample Surveys. *Proceedings of the Social Statistics Section*, American Statistical Association, 242-257.
- Hartley, H.O. et Rao, J.N.K. (1962). Sampling with Unequal Probabilities and without Replacement. *Annals of Mathematical Statistics*, 33, 350-374.
- Hellerman, E. (1982). Overview of the Hellerman I & O Coding System. Ébauche memo, U.S. Bureau of the Census.
- Higginson, J. (1977). *Manual d'utilisation du test de Bell Canada relatif à la sélection de schémas de composition des séries*. Document de recherche n° 77-01-001, Groupe de la désaisonnalisation et des séries chronologiques, Statistique Canada.
- Keyfitz, N. (1957). Estimates of Sampling Variance where Two Units Are Selected from Each Stratum. *Journal of the American Statistical Association*, 52, 503-510.
- Kumar, S. et Lee, H. (1983). Évaluation de l'application d'estimateurs composites à l'enquête sur la population active du Canada. *Techniques d'enquête*, 9, 196-221.
- Lemaître, G.E. (1988). The Measurement and Analysis of Gross Flows. Document de travail No. DMES-88-1E. Statistique Canada.
- Lemaître, G.E. et Dufour, J. (1987). Une méthode intégrée de pondération des personnes et des familles. *Techniques d'enquête*, 13, 211-220.
- Lothian, J. et Morry, M. (1977). The Problem of Aggregation Direct or Indirect Seasonal Adjustment. Document de recherche, Groupe de la désaisonnalisation et des séries chronologiques, Statistique Canada.
- Morry, M. (1975). A Test for Model Selection. Document de recherche n° 75-12-016, Groupe de la désaisonnalisation et des séries chronologiques, Statistique Canada.

Platek, R. and Singh, M.P. (1976). *Methodology of the Canadian Labour Force Survey*. Statistics Canada, Catalogue no. 71-526.

Poterba, J.M. and Summers, L.H. (1986). Report Errors and Labour Market Dynamics. *Econometrica*, 54, 1319-38.

Rao, J.N.K., Hartley, H.O. and Cochran, W.G. (1962). A Simple Procedure of Unequal Probability Sampling without Replacement. *Journal of the Royal Statistical Society*, B, 24, 482-491.

Shiskin, J., Young, A.H. and Musgrave, J.C. (1967). Variant of Census Method II Seasonal Adjustment. Washington, D.C., U.S. Bureau of the Census, Department of Commerce, Technical Paper No. 15.

Singh, A.C. and Lemaitre, G.E. (1989). On an ϵ -Contamination Model for Multinomial Response Error With Application to Labour Flows. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Singh, A.C. and Rao, J.N.K. (1990). Adjustments for Classification Errors in Gross Flows. Presented at the International Conference on Measurement Errors in Surveys, Tucson, Arizona.

Singh, M.P. and Drew, J.D. (1977). Sample Expansion in Self Representing Units of the Canadian Labour Force Survey. Technical Report, Census and Household Survey Methods Division, Statistics Canada.

Tessier, R. (1976). Rotation Group Biases in the Old and New Labour Force Survey. *Survey Methodology*, 2, 256-271.

Platek, R. et Singh, M.P. (1976). *Méthodologie de l'enquête sur la population active du Canada*. Statistique Canada, N° 71-526 au catalogue.

Poterba, J.M. et Summers, L.H. (1986). Report Errors and Labour Market Dynamics. *Econometrica*, 54, 1319-38.

Rao, J.N.K., Hartley, H.O. et Cochran, W.G. (1962). A Simple Procedure of Unequal Probability Sampling without Replacement. *Journal of the Royal Statistical Society*, B, 24, 482-491.

Shiskin, J., Young, A.H. et Musgrave, J.C. (1967). Variant of Census Method II Seasonal Adjustment. Washington, D.C., U.S. Bureau of the Census, Department of Commerce, Technical Paper No. 15.

Singh, A.C. et Lemaitre, G.E. (1989). On an ϵ -Contamination Model for Multinomial Response Error With Application to Labour Flows. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Singh, A.C. et Rao, J.N.K. (1990). Adjustments for Classification Errors in Gross Flows. Présenté à l'International Conference on Measurement Errors in Surveys, Tucson, Arizona.

Singh, M.P. et Drew, J.D. (1977). Sample Expansion in Self Representing Units of the Canadian Labour Force Survey. Document technique, Division des méthodes de recensement et d'enquêtes - ménages, Statistique Canada.

Tessier, R. (1976). Rotation Group Biases in the Old and New Labour Force Survey. *Techniques d'enquête*, 2, 256-271.

No other monthly report on the Canadian Economy has this much to offer

La seule publication à vous offrir autant d'information sur l'économie canadienne

Canadian Economic Observer

The most extensive and timely information source for people who want objective facts and analysis on the Canadian Economy... every month.

Current economic conditions

Brief, "to the point" a current update summary of the economy's performance including trend analyses on employment, output, demand and the leading indicator.

Feature articles

In-depth research on current business and economic issues: business cycles, employment trends, personal savings, business investment plans and corporate concentration.

Statistical summary

Statistical tables, charts and graphs cover national accounts, output, demand, trade, labour and financial markets.

Regional analysis

Provincial breakdowns of key economic indicators.

International overview

Digest of economic performance of Canada's most important trading partners — Europe, Japan and the U.S.

Economic and statistical events

Each month, CEO also publishes a chronology of current events that will affect the economy, and information notes about new products from Statistics Canada.

Consult with an expert

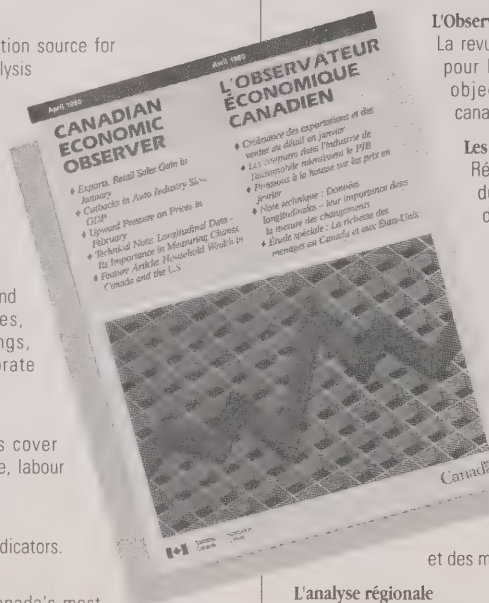
The names and phone numbers of the most appropriate Statistics Canada contacts are provided with each data table in the statistical summary; not only can you read the data and the analysis, you can talk to the experts about it.

The Canadian Economic Observer

(Catalogue no. 11-010) is \$220 annually in Canada, US\$260 in the United States and US\$310 in other countries.

To order, write Publication Sales, Statistics Canada, Ottawa, Ontario, K1A 0T6 or contact the nearest Statistics Canada Regional Reference Centre listed in this publication.

For faster service, fax your order to 1-613-951-1584. Or call toll free at 1-800-267-6677 and use your VISA or MasterCard.



L'Observateur économique canadien

La revue la plus complète et la plus à jour qui soit pour les gens qui désirent des renseignements objectifs et une analyse de l'économie canadienne... chaque mois.

Les conditions économiques actuelles

Résumé bref et incisif de l'actualité économique du mois, comportant l'analyse des tendances de l'emploi, de la production, de la demande et de l'indicateur avancé.

Les études spéciales

Recherche approfondie sur les questions du domaine des affaires et de l'économie : cycles économiques, tendances de l'emploi, épargne personnelle, projets d'investissement et concentration des sociétés.

L'aperçu statistique

Tableaux, graphiques et diagrammes englobant les statistiques des comptes nationaux, de la production, de la demande, du commerce, de l'emploi et des marchés financiers.

L'analyse régionale

Ventilation par province des indicateurs économiques stratégiques.

Le survol de l'économie internationale

Un sommaire du rendement de l'économie des partenaires commerciaux du Canada, comme l'Europe, le Japon et les États-Unis.

Événements économiques et statistiques

Chaque mois, L'OEC publie une chronologie des événements qui influenceront l'économie de même que des renseignements sur les nouveaux produits de Statistique Canada.

Consultez un expert

Les noms et numéros de téléphone des personnes-ressources figurent à chaque tableau de l'aperçu statistique; non seulement pouvez-vous lire les données et l'analyse, mais vous pouvez de plus discuter du sujet avec les experts de Statistique Canada.

L'Observateur économique canadien.

(no 11-010 au catalogue) coûte 220 \$ l'abonnement annuel au Canada, 260 \$ US aux États-Unis et 310 \$ US dans les autres pays.

Pour commander, veuillez écrire à Vente des publications, Statistique Canada, Ottawa (Ontario), K1A 0T6 ou communiquer avec le Centre régional de consultation de Statistique Canada le plus près (voir la liste figurant dans la présente publication).

Pour un service plus rapide, commandez par télécopieur au 1-613-951-1584. Ou composez sans frais le 1-800-267-6677 et utilisez votre carte VISA ou MasterCard.

ORDER FORM		Client Reference Number _____				
Mail to: Fax (613) 951-1584 Publication Sales Statistics Canada Ottawa, Ontario, K1A 0T6 <i>(Please print)</i> Company _____ Department _____ Attention _____ Title _____ Address _____ City _____ Province _____ Country _____ Postal Code _____ Tel. _____		METHOD OF PAYMENT <input type="checkbox"/> Purchase Order Number (please enclose) _____ <input type="checkbox"/> Payment enclosed \$ _____ <input type="checkbox"/> Charge to my: <div style="display: flex; justify-content: space-between;"> <input type="checkbox"/> MASTERCARD <input type="checkbox"/> VISA <input type="checkbox"/> Statistics Canada </div> Account Number _____ Expiry Date _____ <input type="checkbox"/> Bill me later (Max. \$500) Signature _____				
Catalogue No.	Title	Issue	Quantity	Price	Total	
<i>Cheque or money order should be made payable to the Receiver General for Canada/Publications. Canadian Clients pay in Canadian funds. Foreign clients pay in US \$, drawn on a US bank.</i>				PF 03077		
For faster service		1-800-267-6677		MasterCard and Visa accounts <small>Français au verso</small>		

BON DE COMMANDE		Numéro de référence du client _____				
Postez à: Télécopieur: (613) 951-1584 Vente des publications Statistique Canada Ottawa (Ontario), K1A 0T6 <i>(Caractères d'imprimerie s.v.p.)</i> Organisme _____ Service _____ a/s de _____ Fonction _____ Adresse _____ Ville _____ Province _____ Pays _____ Code postal _____ Tél. _____		MODE DE PAIEMENT <input type="checkbox"/> Numéro de la commande (inclure s.v.p.) _____ <input type="checkbox"/> Paiement inclus \$ _____ <input type="checkbox"/> Portez à mon compte: <div style="display: flex; justify-content: space-between;"> <input type="checkbox"/> MASTERCARD <input type="checkbox"/> VISA <input type="checkbox"/> Statistique Canada </div> N° de compte _____ Date d'expiration _____ <input type="checkbox"/> Facturez-moi plus tard (max. 500 \$) Signature _____				
N° au catalogue	Titre	Édition	Quantité	Prix	Total	
<i>Le chèque ou mandat-poste doit être fait à l'ordre du Receveur général du Canada - Publications. Les clients canadiens paient en dollars canadiens; les clients à l'étranger paient en \$ US, tirés sur une banque américaine.</i>				PF 03077		
Pour un service plus rapide, composez		1-800-267-6677		Comptes MasterCard et Visa <small>English on reverse</small>		

BON DE COMMANDE		Numéro de référence du client _____				
Postez à: Vente des publications Statistique Canada Ottawa (Ontario), K1A 0T6 (Caractères d'imprimerie s.v.p.) Organisme _____ Service _____ a/s de _____ Fonction _____ Adresse _____ Ville _____ Province _____ Pays _____ Code postal _____ Tél. _____		MODE DE PAIEMENT <input type="checkbox"/> Numéro de la commande (inclure s.v.p.) _____ <input type="checkbox"/> Paiement inclus _____ \$ <input type="checkbox"/> Portez à mon compte: <div style="display: flex; justify-content: space-between;"> <input type="checkbox"/> MASTERCARD <input type="checkbox"/> VISA <input type="checkbox"/> Statistique Canada </div> N° de compte _____ Date d'expiration _____ <input type="checkbox"/> Facturez-moi plus tard (max. 500 \$) Signature _____				
N° au catalogue	Titre	Édition	Quantité	Prix	Total	
<i>Le chèque ou mandat-poste doit être fait à l'ordre du Receveur général du Canada - Publications. Les clients canadiens paient en dollars canadiens; les clients à l'étranger paient en \$ US, tirés sur une banque américaine.</i>				PF 03077		
Pour un service plus rapide, composez		☎ 1-800-267-6677 ☎			Comptes MasterCard et Visa	

English on reverse

ORDER FORM		Client Reference Number _____				
Mail to: Publication Sales Statistics Canada Ottawa, Ontario, K1A 0T6 (Please print) Company _____ Department _____ Attention _____ Title _____ Address _____ City _____ Province _____ Country _____ Postal Code _____ Tel. _____		METHOD OF PAYMENT <input type="checkbox"/> Purchase Order Number (please enclose) _____ <input type="checkbox"/> Payment enclosed _____ \$ <input type="checkbox"/> Charge to my: <div style="display: flex; justify-content: space-between;"> <input type="checkbox"/> MASTERCARD <input type="checkbox"/> VISA <input type="checkbox"/> Statistics Canada </div> Account Number _____ Expiry Date _____ <input type="checkbox"/> Bill me later (Max. \$500) Signature _____				
Catalogue No.	Title	Issue	Quantity	Price	Total	
<i>Cheque or money order should be made payable to the Receiver General for Canada/Publications. Canadian clients pay in Canadian funds. Foreign clients pay in US \$, drawn on a US bank.</i>				PF 03077		
For faster service		☎ 1-800-267-6677 ☎			MasterCard and Visa accounts	

Français au verso

Introducing

Perspectives on Labour and Income

Canada's essential employment and income information - now in one quarterly journal

If you're responsible for developing employment strategies, negotiating labour contracts, forecasting economic trends, or administering social programs, you'll find ***Perspectives on Labour and Income*** indispensable.

It will keep you up-to-date on the latest Canadian trends in employment... *unemployment insurance, pensions, and industry changes...* and income... *earning gaps between men and women, family income and spending habits*, and more. Every issue of this quarterly journal contains:

- **Feature Articles...** in-depth information on vital topics
- **Forum...** an arena for discussion among researchers and readers
- **Sources...** a compendium of new information sources, news and updates on current research
- **Key Labour and Income Facts...** over 60 indicators let you monitor the trends on a national and provincial level.

Don't miss a single issue. Order your subscription today!

Perspectives on Labour and Income (Catalogue No. 75-001E) is \$53 annually (4 issues) in Canada, US\$64 in the United States and US\$74 in other countries.

To order, write: Publication Sales, Statistics Canada, Ottawa, Ontario, K1A 0T6, or contact your nearest Statistics Canada Regional Reference Centre, listed in this publication.

Nouveau !

L'emploi et le revenu en perspective

Toute l'information essentielle sur l'emploi et le revenu au Canada dans une nouvelle revue trimestrielle

Si vous avez la responsabilité d'élaborer des stratégies d'emploi, de négocier des contrats de travail, de prévoir les nouvelles tendances du marché ou d'administrer des programmes sociaux, vous ne pouvez pas vous passer de ***L'emploi et le revenu en perspective...***

Cette revue vous renseigne sur tout ce qui se passe dans le domaine de l'emploi... *les employés à temps partiel, les pensions, les changements de l'industrie...* et des revenus... *les disparités salariales entre hommes et femmes, le revenu familial et les habitudes de consommation* et plus encore. Chaque numéro de cette revue trimestrielle comprend :

- **Des articles de fond...** des analyses détaillées sur des sujets de l'heure
- **Un forum...** une tribune pour échanger vos idées et connaître l'opinion des autres chercheurs et lecteurs
- **Des sources...** un condensé de nouvelles sources d'information, de renseignements et une mise à jour sur les recherches en cours
- **Des indicateurs clés de l'emploi et du revenu...** plus de 60 indicateurs vous permettant d'analyser les tendances du marché provincial et national.

Ne ratez pas un seul numéro. Abonnez-vous dès aujourd'hui !

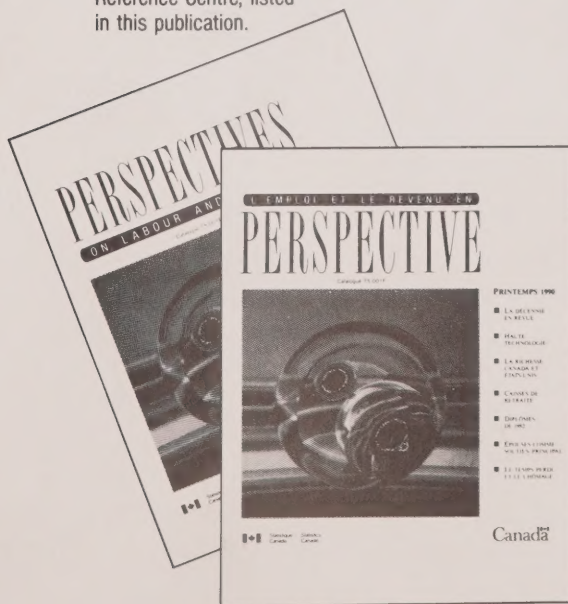
Un abonnement à ***L'emploi et le revenu en perspective*** (n° 75-001F au catalogue) coûte 53 \$ pour quatre numéros par an au Canada, 64 \$ US aux États-Unis et 74 \$ US dans les autres pays.

Pour commander, veuillez écrire à Vente des publications, Statistique Canada, Ottawa (Ontario), K1A 0T6 ou communiquer avec le Centre régional de consultation de Statistique Canada le plus près (voir la liste dans la présente publication).

For faster service, call toll free and use your VISA or MasterCard.

Pour obtenir votre revue plus rapidement, composez le numéro suivant sans frais et portez la commande à votre compte VISA ou MasterCard.

1-800-267-6677



SURVEY METHODOLOGY

*A Journal of Statistical
Development and Applications*

Each article focuses on developing and evaluating specific methodologies for data collection or data evaluation.

In addition to general topics of current interest to survey statisticians, most issues of **Survey Methodology** contain a special section with a concentrated treatment of new techniques and experiences for a selected topic.

All articles in **Survey Methodology** are refereed by an international board and the journal enjoys world wide circulation as a result of cooperative arrangements with various statistical associations.

Invitation to Authors — Authors are invited to submit manuscripts in either English or French. For more information, please write to: Editor, Survey Methodology, Methodology Branch, Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6.

To inquire about availability and prices for Survey Methodology (12-001), in Canada phone 1-800-267-6677. Foreign clients write to Publication Sales, Statistics Canada, Ottawa, Canada, K1A 0T6.

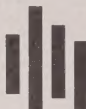
Editorial Board: Editor — M.P. Singh. Associate editors — B. Afonja, D.R. Bellhouse, D. Binder, E.B. Dagum, J.C. Deville, G.D. Drew, W.A. Fuller, J.F. Gentleman, M. Gonzalez, R.M. Groves, D. Holt, G. Kalton, J.N.K. Rao, D.B. Rubin, I. Sande, C.E. Särndal, W. Schaible, F.J. Scheuren, K.M. Wolter. Assistant editors — J. Gambino, L. Mach, A. Thérberge.

Management Board: G.J. Brackstone, N. Chinnappa, G.J.C. Hole, C. Patrick, F. Mayda, R. Platek, D. Roy, M.P. Singh



Survey Methodology publishes articles dealing with various aspects of statistical development such as,

- design issues in the context of practical constraints
- use of different data sources and collection techniques
- total survey error
- survey evaluation
- research in survey methodology
- time series analysis
- seasonal adjustment
- demographic studies
- data integration
- estimation and data analysis methods
- general survey systems development



Techniques d'enquête publie des articles qui portent sur différents aspects des méthodes statistiques :

- les problèmes de conception découlant des contraintes d'ordre pratique
- l'utilisation de différentes sources de données et techniques de collecte
- les erreurs dans les enquêtes
- l'évaluation des enquêtes
- la désaisonnalisation
- les études démographiques
- la recherche sur les méthodes d'enquêtes
- l'analyse des séries chronologiques
- l'intégration des données statistiques
- les méthodes d'estimation et d'analyse de données
- le développement de systèmes généralisés

TECHNIQUES D'ENQUÊTE

*Une revue sur les méthodes statistiques
et leur utilisation.*

Chaque article met l'accent sur l'élaboration et l'évaluation de méthodes particulières de collecte et d'évaluation des données.

En plus de s'attarder aux sujets d'intérêt habituels des statisticiens d'enquêtes, la plupart des numéros de **Techniques d'enquête** contiennent une section traitant en profondeur des nouvelles techniques et expériences concernant un sujet choisi.

Tous les articles de **Techniques d'enquête** sont revus par un comité de rédaction international. D'ailleurs, des ententes coopératives avec différentes associations statistiques internationales assurent à la revue une diffusion mondiale.

Invitation aux auteurs — Les auteurs désirant faire paraître un article sont invités à faire parvenir leur texte, anglais ou français, à : Techniques d'enquête, Rédacteur en chef, Statistique Canada, Ottawa (Ontario), Canada, K1A 0T6.

Pour se renseigner sur la disponibilité et les prix relatifs à Techniques d'enquête (12-001), les clients canadiens peuvent composer le numéro sans frais 1-800-267-6677. Les clients étrangers peuvent écrire à Vente des publications, Statistique Canada, Ottawa, Canada, K1A 0T6.

Comité de rédaction : rédacteur en chef — M.P. Singh; rédacteurs associés — B. Afonja, D.R. Bellhouse, D. Binder, E.B. Dagum, J.C. Deville, G.D. Drew, W.A. Fuller, J.F. Gentleman, M. Gonzalez, R.M. Groves, D. Holt, G. Kalton, J.N.K. Rao, D.B. Rubin, I. Sande, C.E. Särndal, W. Schaible, F.J. Scheuren, K.M. Wolter. **rédacteurs adjoints** — J. Gambino, L. Mach, A. Thérberge.

Comité de direction : G.J. Brackstone, N. Chinnappa, G.J.C. Hole, C. Patrick, F. Mayda, R. Platek, D. Roy, M.P. Singh.